Neuropsychologia 47 (2009) 2995–3006

Contents lists available at ScienceDirect

Neuropsychologia

journal homepage: www.elsevier.com/locate/neuropsychologia

Prefrontal contributions to rule-based and information-integration category learning

David M. Schnyer[a,*], W. Todd Maddox[a], Shawn Ell[b], Sarah Davis[c], Jenni Pacheco[d], Mieke Verfaellie[e]

[a] Department of Psychology, Institute for Neuroscience, University of Texas at Austin, Austin, TX, United States
[b] Department of Psychology, University of Maine, Graduate School of Biomedical Sciences, Orono, ME, United States
[c] Department of Brain and Cognitive Sciences, University of Rochester, Rochester, NY, United States
[d] Department of Psychology, Institute for Neuroscience, University of Texas at Austin, Austin, TX, United States
[e] VA Boston Healthcare System and Boston University School of Medicine, Boston, MA, United States

ARTICLE INFO

Article history:
Received 2 June 2009
Received in revised form 18 July 2009
Accepted 21 July 2009
Available online 28 July 2009

Keywords:
Category learning
Prefrontal cortex
Feedback processing

ABSTRACT

Previous research revealed that the basal ganglia play a critical role in category learning [Ell, S. W., Marchant, N. L., & Ivry, R. B. (2006). Focal putamen lesions impair learning in rule-based, but not information-integration categorization tasks. *Neuropsychologia*, *44*(10), 1737–1751; Maddox, W. T. & Filoteo, J. V. (2007). Modeling visual attention and category learning in amnesiacs, striatal-damaged patients and normal aging. In Advances in Clinical-cognitive science: formal modeling and assessment of processes and symptoms (pp. 113–146). Washington DC: American Psychological Association] but less is known about the specific role of prefrontal cortical (PFC) regions in category learning. The current study examined rule-based (RB) and information-integration (II) category learning in 13 patients with damage primarily to ventral PFC regions. After 600 learning trials with feedback, patients were significantly less accurate than matched controls on both RB and II learning. Model-based analysis identified subgroups of patients whose impaired performance in each task was due to the use of sub-optimal learning strategies. Those patients impaired at either II or RB learning, performed significantly worse on the Wisconsin Card Sorting Test, a test of abstract rule formation and the ability to shift and maintain rules. Lesion analysis pointed to damage in a fairly circumscribed region of ventral medial prefrontal cortex as common to the impaired group of patients and those patients without ventral PFC damage mostly performed normally. These results provide further evidence that the ventromedial prefrontal cortex is critically important for the ability to monitor and integrate feedback in order to select and maintain optimal learning strategies.

1. Introduction

Category learning is a critical cognitive skill that allows us to respond differently to objects and events in different groups (or categories). Cognitive neuroscience has extensively examined the neural substrates of classification learning with computational modeling (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Brown, Bullock, & Grossberg, 1999; Frank, 2005), functional neuroimaging (Filoteo et al., 2005b; Nomura et al., 2007; Poldrack, Prabhakaran, & Gabrieli, 1999; Seger & Cincotta, 2005, 2006), and studies with neurologically damaged patients (Knowlton, Mangels, & Squire, 1996; Maddox & Filoteo, 2005, 2007; Price, 2005). Many of these studies have focused on the contribution of the basal ganglia (BG, e.g., Ashby & Ennis, 2006) to category learning and despite its demonstrated importance, it is clear that the BG is functionally linked to a number of important cortical regions, in particular prefrontal cortex (PFC), that also play important roles (Seger, 2008).

One of the most popular neurobiologically inspired models of category learning is the Competition between Verbal and Implicit Systems model (COVIS; Ashby et al., 1998). COVIS asserts that the long-term learning of different types of category structures is mediated by different systems that have unique neural substrates (Ashby & Ell, 2001; Ashby & Maddox, 2005; Maddox & Ashby, 2004). Two types of category structures that have been studied extensively include rule-based (RB) and information-integration (II) category structures. Rule-based tasks are those in which learning is thought to rely on a frontally mediated hypothesis-testing system interacting with portions of the BG, namely the anterior caudate. In general, the rule that maximizes accuracy can be verbally described. One classic rule-based categorization task, the Wisconsin Card Sorting Test (WCST; Milner, 1963), has been shown to be highly dependent on the PFC (Lombardi et al., 1999; Monchi, Petrides, Petre, Worsley, & Dagher, 2001) and it is commonly used in neuropsychological assessment to test for PFC damage or dysfunction (Milner, 1963). In addition to neuropsychological evidence, neuroimaging work has

* Corresponding author.
E-mail address: schnyer@psy.utexas.edu (D.M. Schnyer).

0028-3932/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neuropsychologia.2009.07.011

also supports the role of PFC in RB category learning (Aron et al., 2004; Filoteo et al., 2005b; Monchi et al., 2001; Rao et al., 1997; Seger & Cincotta, 2006).

In contrast to RB tasks, II tasks are those in which accuracy is maximized when information from two or more stimulus dimensions is integrated at some pre-decisional stage. With II categories, the rule that maximizes accuracy cannot be described verbally, and learning involves incremental acquisition of stimulus–response associations (Ashby & Waldron, 1999). It has been proposed that learning of II categories is mediated by a procedural learning system involving the BG in interaction with posterior perceptual regions and recent fMRI evidence supports this characterization (Nomura et al., 2007; Seger & Cincotta, 2002, 2005). Unlike RB learning, II learning is assumed not to rely on the PFC; experimental evidence appears to support this as II learning is unaffected by dual task conditions that put increased burden on executive processes, thought to be mediated by PFC (Maddox, Ashby, Ing, & Pickering, 2004a). In contrast, RB learning is affected by dual task conditions. More recent evidence indicates that decreasing the influence of executive processes can actually increase the type of implicit learning that is likely engaged in II categorization (Filoteo, Lauritzen, & Maddox, in press). These findings support the proposition that II learning is less dependent on processes mediated by PFC than is RB learning.

Given that the COVIS framework clearly specifies that RB learning relies on a frontally mediated hypothesis-testing system and in this way differs from II learning in the need for direct PFC engagement, a straightforward prediction might be that damage to PFC would impair RB learning but leave II learning intact. To date, evidence for such a neurological dissociation is sparse. One of the reasons for this may be that despite the differences between the neurobiology of the II and RB category learning systems, both types of learning rely on effective feedback processing (Maddox & Ashby, 2004). Research clearly indicates that response to feedback relies, at least in part, on specific regions of the PFC. One of the regions that has been critically implicated in feedback-based learning is the ventral PFC (vPFC; Cools, Clark, Owen, & Robbins, 2002; Fellows & Farah, 2003; Haber, Kim, Mailly, & Calzavara, 2006). Ventral PFC has been tied to learning both when feedback reflects information about expected outcomes (Takahashi et al., 2009) and when expectations are violated (Monchi et al., 2001; Takahashi et al., 2009) and it has been shown to be critical to induce strategy shifts in order to optimize performance (Ghods-Sharifi, Haluk, & Floresco, 2008). Given the critical role of vPFC in feedback learning, it would be reasonable to hypothesize a general role of vPFC in both RB and II category learning.

Feedback in category learning serves at least two basic purposes. First, it serves to facilitate dopamine-mediated reinforcement learning in the corticostriatal circuits (Ashby, Ennis, & Spiering, 2007) where unexpected rewards trigger dopamine release (Schultz, 1998) and thereby strengthen the synaptic connections that support visuomotor learning. Second, feedback provides a signal (particularly following negative feedback) that helps guide rule selection and strategy shifting (Monchi et al., 2004; Seger, 2008). It is likely that strategy shifting that results from feedback is one common component in both RB and II category learning. For instance, within rule-based learning, optimal performance requires trying different types of rule structures (i.e. unidimensional versus conjunctive rules) and dropping ineffective approaches in response to feedback (Lawrence, 2000). This type of strategy shifting is similar to what is involved in the Wisconsin Card Sorting Test (WCST), where surprising negative feedback forces a person to drop the previously learned categorization rule (e.g., on shape) and to shift to trying new ones (e.g., on color). By contrast, in II learning, optimizing performance requires a general shift from a rule-based approach to the more automatic information-integration strategy (Ashby et al., 1998).

It is currently unknown whether these two types of strategy shifts in response to feedback, one within the class of RB strategies and one involving a shift from RB to II strategies, will depend on the same region of vPFC. More generally, it is also currently unknown whether RB and II category learning both rely on a common region of the PFC, despite the emphasis on greater PFC reliance for RB learning relative to II learning. To examine this question, the current study examined category learning in individuals with lesions primarily to ventral PFC regions. Both RB and II category structures were examined to assess whether the vPFC plays a similar role in feedback learning across different types of category learning. Given the role of ventral PFC in altering strategy in response to feedback, it was predicted that persons with vPFC lesions would have difficulty shifting between different classes of strategy (shifting from RB to II) as well as shifting within a single class of strategies (RB) in response to feedback. Despite the extensive literature on reinforcement learning, we are unaware of any studies that have directly examined this issue, comparing RB and II category learning in patients with frontal lesions.

2. Experiment

The stimuli were lines that varied in length and orientation, assigned to one of four categories. A scatterplot of the stimuli along with the optimal decision bounds are displayed in Fig. 1. Because the rule-based and information-integration tasks are related via a simple rotation, the two tasks are equated on task difficulty, optimal accuracy, and the number of relevant dimensions (Ell, Marchant, & Ivry, 2006; Maddox, Filoteo, Hejl, & Ing, 2004b). Importantly, in both tasks participants must attend equally to both length and orientation to maximize accuracy. In the rule-based task, accuracy is maximized when the participant adopts a conjunctive strategy that involves deciding if the line is long or short and if the angle is steep or shallow, and then combining those decisions. The optimal strategy is to respond "A" to short, shallow angle lines, "B" to short, steep angle lines, "C" to long, shallow angle lines, and "D" to long, steep angle lines. Thus, the dimensional integration is post-decisional, verbalizable, and involves a rule-based strategy because participants must first make a decision about each dimension (e.g. length: short and orientation: shallow) and then combine those two decisions to form a conjunction (i.e. short and shallow implies category A; Ashby & Gott, 1988; Shaw, 1982).

For the information-integration task, the categories were created by rotating the rule-based categories 45° counterclockwise. Accuracy maximization again involves the integration of length and orientation information, but in this case the integration is pre-decisional. It is important to note that reasonable performance levels can be achieved by applying verbal rule-based strategies in the information-integration task. In fact, rule-based strategies are often used early in training with information-integration tasks. We applied quantitative modeling techniques at the individual participant level to determine the type of strategy (rule-based or information-integration) used by each participant in each condition (details presented below).

3. Methods

3.1. Participants

Thirteen patients with damage to frontal cortex (9 females and 4 males) who were native speakers of English completed the task (age range 49–78 years). Patients were referred to the Memory Disorders Research Center for evaluation as a result of complaints about cognitive functioning. The frontal patients all had focal lesions secondary to stroke, aneurysm, or trauma and were in stable neurological condition at the time of testing (see Table 1a for demographic information, damage etiology and lesion location descriptions). None of the patients showed significant language impairment that would have interfered with task performance. CT and/or MRI were available for all patients. Using these scans, the site of lesion

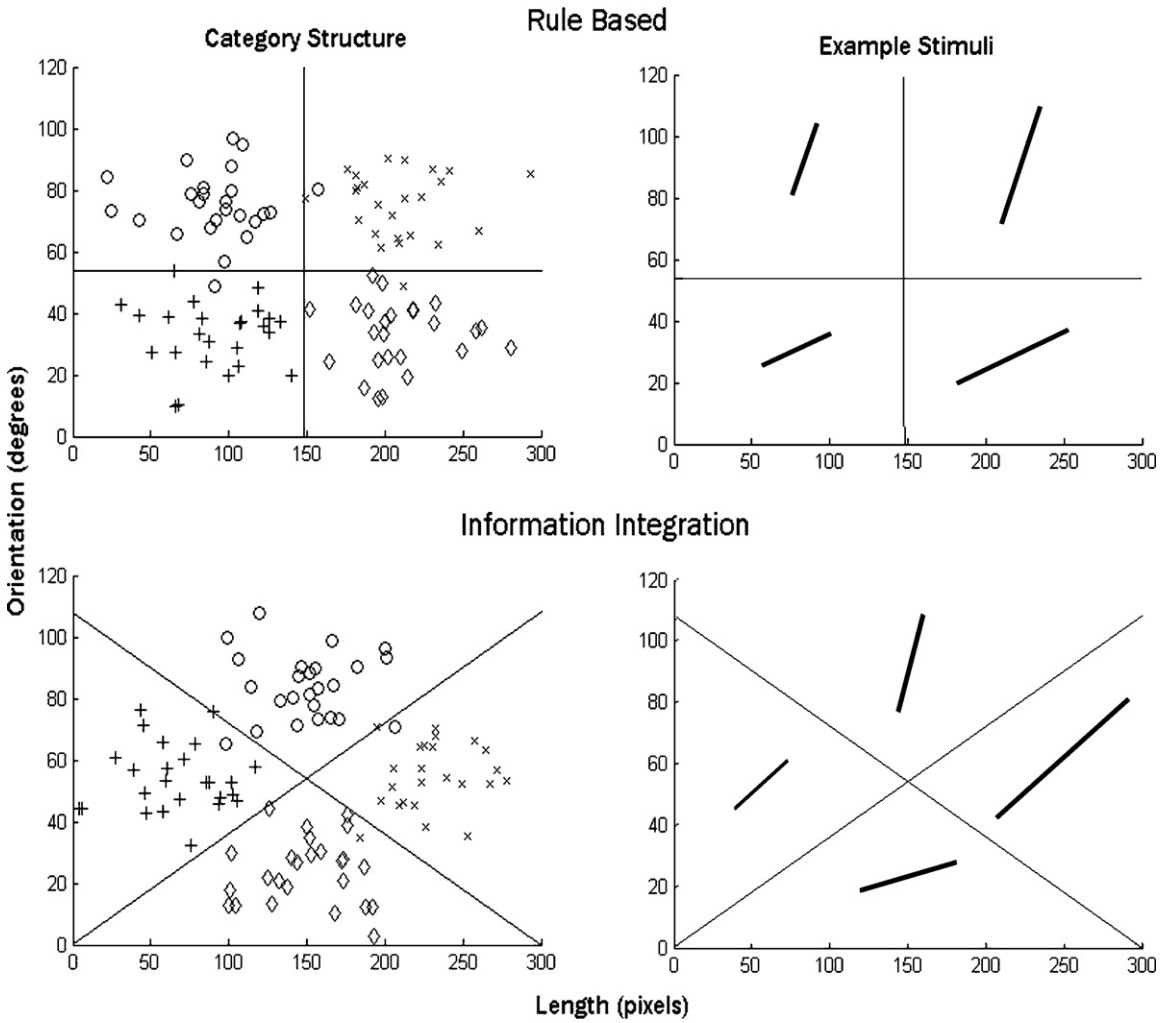

Fig. 1. The category structures for both the RB and II tasks. Scatterplot of the stimuli in length-orientation space in the two tasks (left panels) along with example stimuli (right panels). Each point in the scatterplot represents a single stimulus. Category 1 exemplars are plotted as plus signs, Category 2 exemplars as circles, Category 3 exemplars as diamonds, and Category 4 as symbol '×'. The solid lines are the optimal decision boundaries. Copyright © 2006 by Elsevier. Reproduced with permission from Ell et al. (2006).

was identified by projection on a standard brain oriented in MNI space. Determination of whether a patient had damage to regions of the BG was performed by a neurologist. Only one patient had damage extending outside the frontal cortex.

Control participants consisted of 7 females and 4 males who were matched to patients in age and WAIS-III verbal IQ (see Table 1, F's < 1). These individuals underwent appropriate health screening, had normal or corrected to normal vision, and were free of past or current neurological disorders or psychiatric disability.

Table 1a
Demographic, etiology and lesion location information for each of the 13 PFC patients and the mean values for controls.

	Gender	Age at testing	Verbal IQ (scaled)	Etiology	Lesion location	
					Overall	Basal ganglia
FL_01	f	49	101	Infarct	Left-DL and VM frontal WM; right-DM	No
FL_02	f	60	124	Aneurysm clip	Right-VM	No
FL_03	f	73	88	Infarct	Bilateral-VM	No
FL_04	m	55	117	Infarct	Right-caudate, VM, right polar frontal	BG (rt)
FL_05	f	68	103	Infarct	Bilateral-VM and DM	No
FL_06	f	60	105	Aneurysm clip	Right-DM, polar frontal	No
FL_07	m	63	137	Hematoma/contusion	Right-ventral frontal/temporal	No
FL_08	f	63	114	Hemorrhage	Left-DL	No
FL_09	f	78	98	Infarct	Bilateral-VM; left-DM	No
FL_10	f	65	93	Infarct	Bilateral-VM	No
FL_11	m	77	82	Hemorrhage	Left-DL, putamen	BG (lft)
FL_12	f	64	93	Infarct	Bilateral-VM, CC; right-caudate	BG (rt)
FL 13	m	62	99	Infarct	Bilateral-VM; left-caudate	BG (lft)
Patients	4 male; 9 female	64.4	104.2			
Controls	4 male; 7 female	62.7	109.8			

CC = corpus callosum; DL = dorsal lateral; DM = dorsal medial; WM = white matter; VM = ventral medial; lft = left, rt = right.

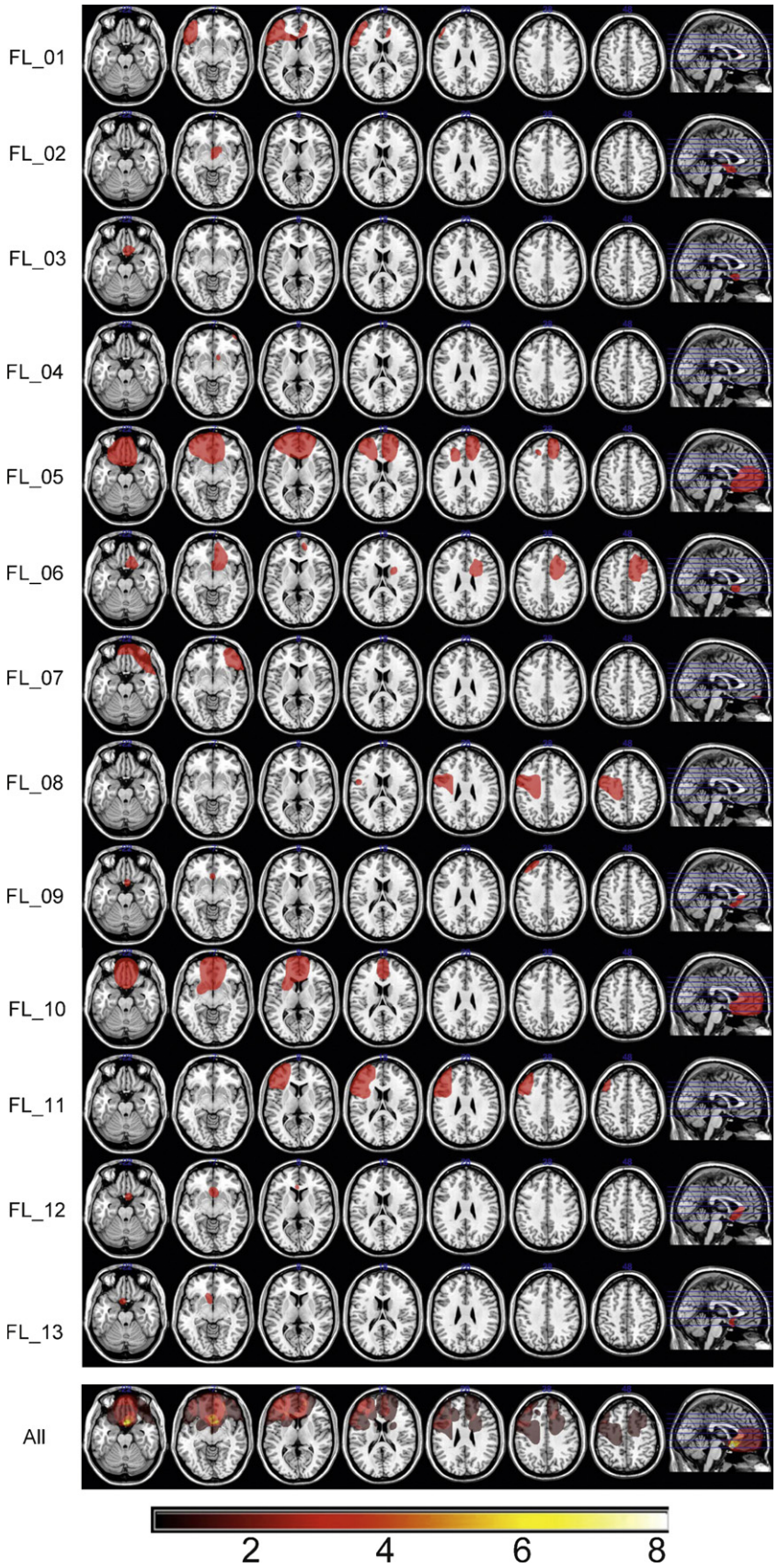

Fig. 2. Patient lesions displayed on a common atlas template (MNI151). Lesions were hand drawn from clinical or experimental MRI scans or clinical CT. They represent an approximation of the region of brain damage and often "overestimate" the extent of this damage. The final row shows the lesion overlap across all 13 patients and a color scale indicating the extent of lesion overlap. Scans are in neurological format, right = right.

Written informed consent was obtained from each volunteer prior to the session. The Human Subjects Committees of Boston University School of Medicine, the Department of Veteran Affairs Medical Center and the University of Texas approved all procedures and all participants were remunerated $20 for their participation.

3.2. Standardized neuropsychological testing

For the purpose of matching general verbal intelligence across controls and patients, each participant completed the Vocabulary subtest of the Wechsler Adult Intelligence Scale, Third Edition (WAIS-III, Wechsler, 1997). This subtest is accepted to be a good measure of both verbal and general mental ability (Lezak, 1995). In order to investigate the relationship between the experimental measures of category learning and measures of memory and executive functioning, frontal patients were administered a selection of standardized neuropsychological tests. Memory performance was evaluated using the Wechsler Memory Scale, Third Edition (WMS-III; Wechsler, 1997) and the Warrington Recognition Memory Test (WRMT) for words (Warrington, 1984). In the case of the WMS-III, the General Memory (GM) score was used as a composite measure of new episodic learning ability. This is comprised of the delayed subtest scores from Logical Memory, Verbal Paired Associates, Facial recognition, and Family Pictures. The WRMT for words provides a measure of recognition memory in the verbal domain while the GM of the WMS-III provides a measure of recognition across multiple memory domains. Both of these measures are sensitive to the types of memory dysfunction often seen in patients with damage to frontal cortex.

For assessment of executive functions, frontal patients completed tests of word generation (FAS; Stuss & Benson, 1986), complex visual scanning and tracking (Trails B; Partington & Leiter, 1949), category formation and set shifting (WCST; Nelson, 1976), and salient response inhibition (Stroop Color–Word Test; Stroop, 1935). In the Controlled Oral Word Association (FAS) Test, subjects are asked to generate as many different words as possible that begin with a particular letter during a 1 min period and the total number of words generated for all three letters comprises the verbal fluency score. The Trail Making Test–Part B is a visual conceptual and visuomotor tracking task that requires connecting consecutively numbered and lettered circles on a paper work sheet by alternating between the two sequences. Trails B is sensitive to frontal-lobe dysfunction, and it has been proposed that performance is indicative of the subject's ability to shift set and process concurrent stimuli (Lezak, 1995). The Wisconsin Card Sorting Test assesses a person's ability to form abstract concepts, utilize feedback, and to shift and maintain set. Scores reflect both the ability to move through cards in an effective manner (total concepts obtained) as well as the ability to disengage from previous concepts (perseverative errors). The Stroop Color–Word Test (Stroop, 1935) measures the ability to inhibit inappropriate responses in the presence of interfering stimuli (Lezak, 1995).

3.3. Lesion analysis

To examine the relationship of lesion site to RB and II category learning in the frontal patients, lesion information was extracted from clinical CT or T1 MRI scans and drawn on a brain oriented in standard MNI305 space using MRIcro (www.mricro.com). Convergence of lesions was observed by retaining overlapping voxels and projecting them onto a "generic" brain oriented in the standardized MNI space. Lesion location templates for all frontal patients are illustrated on axially oriented slices (Fig. 2). Examination of function–lesion relationships was accomplished by characterization of the lesion overlap in patients whose performance in both of the two category learning tasks was impaired.

3.4. Stimuli and stimulus generation

The 100 stimuli used in the RB task (25 from each category) and 100 stimuli used in the II task (25 from each category) were generated in a manner identical to those used by Ell et al. (2006), Maddox, Ashby, et al. (2004) and Maddox, Filoteo, et al. (2004) (see Fig. 1). The category distribution parameters used to generate the stimuli were taken from Maddox, Ashby, et al. (2004) and Maddox, Filoteo, et al. (2004) where each category was defined as a bivariate normal distribution with a mean and a variance on each dimension, and by a covariance between dimensions. For the rule-based task, twenty-five random samples (x, y) were drawn from each category distribution, and each sample was used to construct a single line with some length (in pixels) and orientation that was converted to radians by multiplying the sample value by $\pi/500$. The scale factor ($\pi/500$) was selected based upon past research in an effort to equate the discriminability of changes in perceived length to changes in perceived orientation. A linear transformation was performed to ensure that the sample and population means, variances, and covariances were identical. The stimuli used in the information-integration task were generated by rotating each of the rule-based stimuli 45° clockwise around a central point located at 150 pixels in length (4° of visual angle) and 150 orientation units (i.e., 54° from horizontal). The order of the 100 stimuli was randomized separately for each block and each participant. Each stimulus was presented on a black background and subtended a visual angle ranging from 0.7° to 7.3° at a viewing distance of approximately 60 cm. The stimuli were generated and presented using the Psychophysics Toolbox extensions for MATLAB (Brainard, 1997; Pelli, 1997). The stimuli were displayed on a laptop LCD with 1024 pixel × 768 pixel resolution.

4. Procedure

Participants were tested in two sessions separated by at least one week. Assignment of task (RB or II learning) to session was counterbalanced between subjects. Sessions began with a short introduction to the task in which the participant was told that he/she would see a single stimulus on each trial and was to make a category assignment by pressing one of four response keys with either index finger. After responding, feedback regarding the correctness of the response (correct: green cross; incorrect: red cross) along with the correct category label was presented in the center of the screen for 1 s. The screen was then blanked for 500 ms prior to the appearance of the next stimulus. In addition to trial-by-trial feedback, feedback was given at the end of each block of 100 trials regarding the participant's accuracy during that block. The participant was told that there were four equally likely categories and was informed that the best possible accuracy was 95% (i.e., optimal accuracy). In addition, he/she was told that there was no response time limit. The laptop keyboard was used to collect responses with the characters 'z', 'w', '/', and 'p' assigned to categories 1–4, respectively. Following Maddox, Ashby, et al. (2004) and Maddox, Filoteo, et al. (2004), the category numbers did not appear on the response keys and the response mappings were fixed across participants, however, great care was taken to instruct the participants as to the category–response key mappings.

Each participant completed six test blocks of 100 trials for each task. Within each block, the ordering of the 100 stimuli was randomized between subjects and throughout the first block the experimenter repeated the instructions as needed and provided encouragement. When necessary, the experimenter reminded the participants of the category–response key mappings during the first block. After the first block, there were only brief breaks between the remaining blocks.

5. Results

5.1. Category learning accuracy

For both category learning structures, mean accuracy was calculated for each block on a participant by participant basis. Graphs of the group averages across the six blocks are presented in Fig. 3a and b for RB and II structures, respectively. These accuracy values across both types of category learning were examined in a $2 \times 2 \times 6$ repeated measures ANOVA with group (patients, controls) as a between subjects factor and category structure (RB, II) and block (1–6) as within subjects factors. There was a main effect of group, indicating that patients were less accurate overall than controls ($F[1,22] = 5.42$, $p < 0.05$). There was also a main effect of category structure ($F[1,22] = 15.04$, $p < 0.01$) that did not interact with group or block indicating that participants were more accurate at II structures (mean = 0.622, SE = 0.032) than they were at RB structures (mean = 0.512, SE = 0.031). A main effect of block indicated the expected learning curve with repeated learning trials ($F[5,110] = 57.60$, $p < 0.001$). This effect did not differ between category structures or groups. Interactions between task and block as well as task, block and group were non-significant ($F[5,110] < 1.7$). In short, patients with lesions to PFC demonstrated both a RB and II learning deficit and this was consistent across all six learning blocks.

5.2. Demographic and neuropsychological predictors of patient learning

To examine the relationship between category learning and demographic and neuropsychological measures, correlations were calculated with the scores in the final block for both II and RB

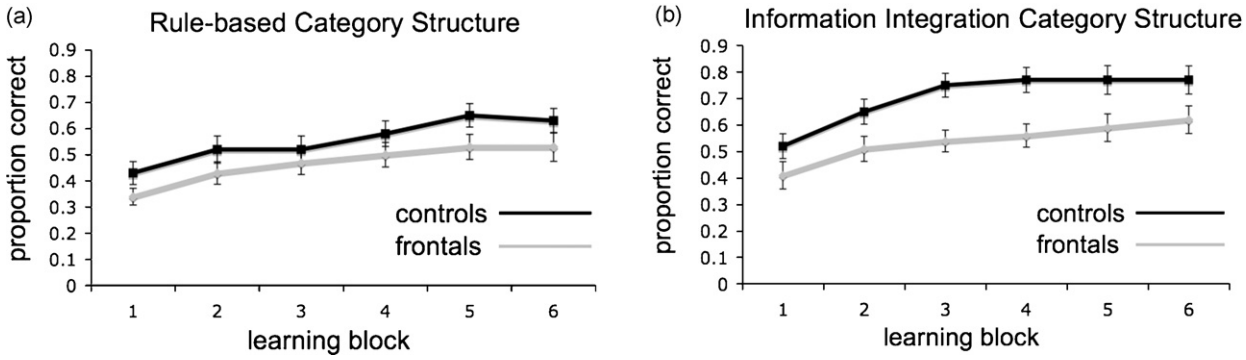

Fig. 3. Task accuracy for RB (panel a) and II (panel b) category structures for both patients and controls across the six learning blocks. Error bars are standard error of the mean.

structures (see Table 1b for neuropsychological measures and final block performance for each patient). Correlations with age and VIQ were examined across the entire group of patients and controls. Age was not significantly correlated with either RB ($r = -0.18$) or II ($r = -0.19$) block 6 performances, whereas Verbal IQ was significantly correlated with both RB ($r = 0.66$, $p < 0.01$) and with II performance ($r = 0.47$, $p < 0.05$).

Correlations with neuropsychological measures were restricted to the patient group and used age-corrected scaled scores (GM) or *Z* scores. A single composite score was created for the WCST from three performance measures (total categories, total errors and perseverative errors) and for the Stroop task from two measures (total time to read the printed color of color words and a reading/naming normalized score reflecting resistance to interference). In the domain of memory, there was a trend towards significant correlations between both RB and II performance and the General Memory scores (RB, $r = 0.53$, $p < 0.10$; II, $r = 0.54$, $p < 0.09$). There was also a nearly significant relationship between RB performance and verbal recognition memory scores (WRMT, $r = 0.52$, $p < 0.06$) but no significant relationship between II performance and WRMT ($r = 0.36$, $p > 0.20$).

Within the domain of executive function, the WCST composite score was significantly correlated with the last block of RB performance ($r = 0.72$, $p < 0.001$), but not with II performance ($r = 0.44$, $p < 0.14$). The opposite pattern was seen for correlations with the Stroop composite: a non-significant relationship with RB performance ($r = 0.50$, $p < 0.13$) and a significant correlation with II performance ($r = 0.60$, $p < 0.05$). Of the remaining executive measures (FAS, and Trails B) there was only a trend towards a correlation between FAS and RB performance ($r = 0.49$, $p < 0.09$).

5.3. Model-based analysis

The accuracy-based analyses suggest that frontal patients were impaired at II and RB category learning. In this section, we gain a more detailed understanding of the locus of the RB and II deficit by applying a series of decision bound models to the data (Ashby & Maddox, 1993; Maddox & Ashby, 1993). Because of concerns with modeling aggregate data, all models were fit separately to the final block of data from each participant (e.g., Estes, 1956; Maddox, 1999; Smith & Minda, 1998). Every model consists of a set of decision bounds that partition the stimulus space into separate response regions. For example, one rule-based model might classify lines as short or long depending upon whether they are less than or greater than 150 pixels, and might assign lines as shallow or steep depending upon whether they are less than or greater than 45°. These decisions are then combined to generate categorization responses with short, shallow lines being assigned to category A, short, steep lines to category B, long, shallow lines to category C, and long, steep lines to category D. This model would be applied to a set of data and a measure of "fit" is computed. Although somewhat more complex, the measure of fit is similar to computing the proportion of the participant's responses that match with the model's response. The model fitting algorithm would then adjust the pixel value used to separate short from long lines, and would adjust the orientation value used to separate shallow from steep lines until the pixel value

Table 1b
Task performance and neuropsychological measures for the 13 patients.

	Block 6 model fits		Block 6 proportion correct		WMS General Memory (scaled)	Warrington words (*Z*-score)	WCST composite (*Z*-score)	Stroop composite (*Z*-score)	FAS (*Z*-score)	Trails B (*Z*-score)
	RB	II	RB	II						
FL_01	RR	NO	0.26	0.5	107	1.38	−0.05	−0.05	−0.13	0.58
FL_02	RB	RR	0.62	0.27	82	−0.22	0.22	–	1.14	−0.89
FL_03	NO	NO	0.49	0.45	67	−2.44	−1.07	−1.07	−0.03	2.61
FL_04	RB	II	0.74	0.76	–	1.11	0.59	0.59	1.79	−1.16
FL_05	NO	NO	0.53	0.72	111	−0.22	−0.78	−0.78	2.50	0.46
FL_06	RB	II	0.75	0.86	86	0.89	0.84	0.84	1.14	−0.37
FL_07	RB	II	0.88	0.9	114	1.56	1.12	1.12	1.86	0.33
FL_08	RB	II	0.68	0.81	104	1.11	1.08	1.08	−0.03	1.53
FL_09	NO	NO	0.49	0.73	73	−2.89	−0.65	−0.65	1.41	−0.84
FL_10	RR	NO	0.32	0.42	60	−0.22	−0.12	–	0.78	1.27
FL_11	NO	II	0.36	0.56	–	−4.44	−0.81	−0.81	−2.19	2.48
FL_12	RR	NO	0.25	0.51	54	−2.44	−3.73	−3.73	0.33	0.41
FL 13	NO	II	0.49	0.61	77	0.67	−0.51	−0.51	1.86	0.05
Patients			0.53	0.62						
Controls	RB	II	0.63	0.77						

RR = random response; NO = non-optimal; II = information-integration; RB = rule-based.

and orientation value that maximize the correspondence between the participant's and the model's responses was achieved.

Different models make different assumptions about the type of strategy that the participant is using. The models allow us to determine whether each participant is using the task-appropriate strategy or a non-optimal strategy to solve the task. One class of models is compatible with the assumption that participants used an explicit hypothesis-testing strategy (RB), a second class is consistent with the assumption that participants used an implicit procedural-based learning strategy (II), and a third class of models is consistent with the assumption that participants were guessing (i.e., responded randomly, RR). When a participant's approach to a task was best fit by the strategy that did not match the task (i.e. using an RB approach in an II task or vice versa), then this is termed a non-optimal strategy. The details of each model and the model fitting procedure are outlined in Appendix A.

By the final block of trials all control subjects had adopted the appropriate strategy—that is an RB strategy in the RB task, and an II strategy in the II task. By contrast, in the RB task, five patients had adopted a RB approach, five a non-optimal approach (i.e., II) and three were responding randomly (RR). In the II task, six patients adopted an II strategy, six a non-optimal strategy (i.e., RB) and one responded randomly. The strategy adopted by the 6th block for each patient as well as performance levels for this block can be seen in the first columns of Table 1b.

Sixth block accuracy was examined as a function of the particular strategy that was adopted. Patients responding randomly and those using a non-optimal strategy were classified together and contrasted against those patients who adopted the optimal strategy for each task. For both category structures, accuracy was significantly higher for patients who adopted the task appropriate strategy (i.e., RB in the RB task and II in the II task) than for patients who adopted a non-optimal strategy or responded randomly (RB, $F[1,12] = 29.53$, $p < 0.001$; II, $F[1,12] = 7.72$, $p < 0.02$, see Fig. 4). Interestingly, even when the random responders are excluded and only those using a non-optimal strategy are compared to patients using an optimal strategy, the results are the same (RB, $F[1,9] = 25.24$, $p < 0.01$; II, $F[1,11] = 6.13$, $p < 0.03$). Furthermore, when the performance of patients who adopted the task appropriate strategy was compared directly with that of controls (who all adopted the task appropriate strategy), there were no group differences in either task (RB, $F[1,16] = 0.071$, ns; II, $F[1,15] = 1.69$, ns, see Fig. 4).

Finally, when neuropsychological test performance was examined as a function of strategy employed by patients in each task, it was primarily the WCST performance that differentiated the two subgroups across both RB and II tasks. Using one-way ANOVAs to examine neuropsychological performance differences between the optimal and non-optimal plus random responder subgroups, there was a significant difference in WCST performance ($F[1,12] = 10.05$, $p < .01$) for the RB task as well as a trend for the II task ($F[1,12] = 4.02$, $p < 0.07$). Again, when random responders are excluded the WCST separates the two groups even more clearly ($F[1,9] = 64.67$, $p < 0.001$) for the RB task as well as for the II task ($F[1,11] = 4.95$, $p < 0.05$). No other frontal measures revealed significant differences between groups in either the RB or II task ($p > 0.16$). One measure of verbal memory (WRMT) and verbal IQ (VIQ) differentiated patients who learned the optimal strategy versus non-optimal and random responders in the RB task only (WRMT, $F[1,12] = 5.51$, $p < 0.05$; VIQ, $F[1,12] = 29.53$, $p < 0.001$).

The model-based analysis demonstrated that those patients who adopted a non-optimal strategy regardless of the category structure and those who were responding randomly accounted for the patient group's impaired performance relative to matched controls. In addition, the group of impaired patients also performed significantly worse on the WCST. Finally, the only apparent difference in neuropsychological performance between the two categorization tasks revealed through the model-based analysis was a relationship between impairment on the RB task and verbal learning and verbal IQ—with impaired patients showing a significantly lower level of performance on the WRMT and the VIQ.

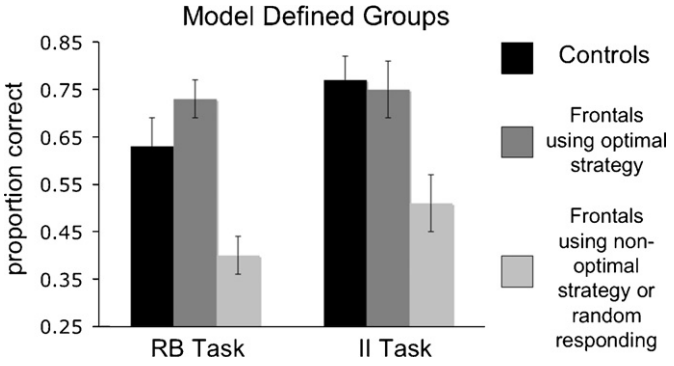

Fig. 4. Task accuracy for both RB and II category structures with patients divided on the use of optimal or non-optimal strategies for the task. Error bars are standard errors of the mean.

5.4. Lesion-based analysis

The relationship between specific cortical damage and impaired category learning was examined by creating a lesion overlap image for the patients who adopted either a random response or non-optimal strategy for both tasks, since these were the only patients to yield impaired performance. Three patients were impaired on only one task and not the other (FL_02, FL_11 and FL_13) so they were initially excluded from the lesion analysis. The group impaired at both tasks included six patients (FL_01, FL_03, FL_05, FL_09, FL_10, and FL_12). For this group, the common region of lesion overlap was a fairly circumscribed region of ventral medial prefrontal cortex (VMPFC; see Fig. 5), where all six patients had damage. Of the three patients who were only impaired on one of the tasks, FL_02 and FL_13 were impaired at II learning and both had small lesions in VMPFC, while FL_11 was impaired only at RB and had a lesion in left dorsal lateral PFC (DLPFC). Therefore, of the patients who were impaired at one or both category learning tasks, nine had lesions to VMPFC and one had a lesion to left DLPFC. Consistent with the neuropsychological results, the lesion analysis points to a common lesion site associated with the inability to adopt the optimal strategy for the specific category structure.

With only a small group of patients without vPFC damage, it is difficult to draw definitive conclusions about the specificity of the lesion-site/impairment results. However, the performance of patients FL_4, FL_8, and FL_11, all of whom lack damage to vPFC, is suggestive. Two of these patients adopted the appropriate strategy by block 6 for both the RB and II task (FL_4 and FL_8), while patient FL_11 was impaired only for the RB task, having adopted the appropriate strategy for the II task. Finally, one patient who adopted appropriate strategies for both tasks had a lesion that included damage to vPFC, but it was unilateral in the right hemisphere (FL_6).

Four (31%) of the 13 frontal patients in this study had some damage to portions of the basal ganglia in addition to their cortical damage as determined by clinical radiological assessment (see Table 1). Given the demonstrated role of the BG in category learning, it would be important to understand whether the BG damage contributed to impaired task performance. In this regard, the BG damage did not appear to be specific to the identified impairments in category learning. Of the four clearly unimpaired patients, one had BG damage (25%). Moreover, of the six patients impaired at both tasks, only one had BG damage (17%). Finally, of the three patients impaired at only one task, two had BG damage (66%) and

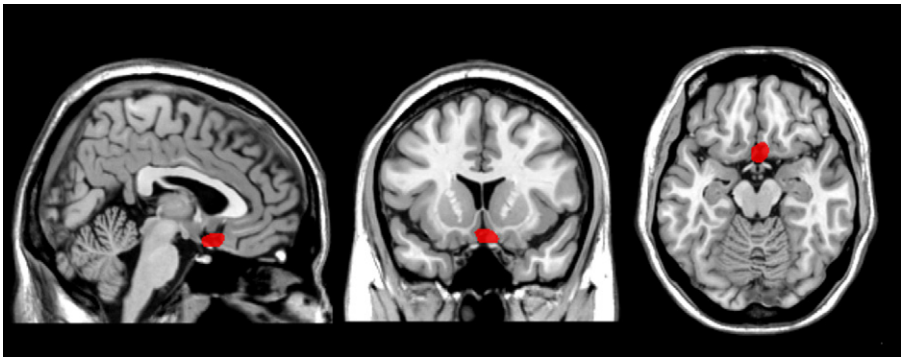

Fig. 5. Lesion overlap projected on a standard brain atlas (MNI151) for the six patients adopting non-optimal or random response strategies in both RB and II tasks. The region in red represents the overlap of the six patients and all impaired patients had damage to VMPFC.(For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

these two were both impaired at RB learning. While not conclusive, these results indicate that BG damage was likely orthogonal to the impairment captured in the model-based analysis.

5.5. General discussion

The current study examined category learning utilizing two different category structures, rule-based and information-integration, in a group of patients with a heterogeneous distribution of damage to frontal cortex. For both category structures, overall learning was impaired relative to age and VIQ matched controls. Model-based analysis identified a subgroup of patients in each learning task who were responsible for the group impairment. These were patients who by the last 100 trials of a 600 trial learning session were unable to adopt and/or maintain the task appropriate strategy. Moreover, patients impaired at either task also performed significantly worse on the WCST than their non-impaired counterparts. Finally, examining the lesion overlap of the impaired patients indicated a common region of VMPFC that was damaged for both those patients impaired at II learning and those impaired at RB learning.

Rule-based and II learning have been postulated to involve two different learning systems mediated by separate neural circuits (Ashby et al., 1998). The system implemented in rule-based learning utilizes explicit, verbally describable rules. It has been proposed that regions of PFC interacting with the anterior caudate mediate this system. By contrast, the II system is an implicit learning system mediated primarily by the posterior caudate in conjunction with posterior perceptual regions. COVIS (Ashby et al., 1998) predicts that depending on the category structure of the task, one or the other will eventually come to dominate performance. While this framework predicts a dissociation with respect to the effects of PFC lesions on performance, both tasks rely on the effective use of feedback and we predicted that both would be susceptible to damage in regions of vPFC that have been shown to be critical to feedback-based learning (Cools et al., 2002; Fellows & Farah, 2003; Haber et al., 2006). As predicted, patients with damage to VMPFC were equivalently impaired in both II and RB across all six learning blocks. Finally, although both RB and II learning involve some portion of the BG, BG damage in a subgroup of the patients tested here could not account for the deficits in performance.

Effective processing of feedback in category learning is essential for testing different verbal rules (as in a rule-based task), and for knowing when or whether to switch strategies to improve performance (as in an information-integration task). We address each of these in turn. The rule-based task used in the current study is complex and involves learning the verbal rules that determine category assignment for four separate categories. Feedback processing is critical in this case and a deficit in this ability may lead to the use of a non-optimal strategy, where only a subset of the categories is actually learned. Interestingly, a careful examination of the response patterns for the patients who used non-optimal strategies in the rule-based task suggests that in all cases, impaired patients were unable to learn all four categories with a reasonable measure of success. Of the five impaired patients, two performed below 25% accuracy for one of the four categories, one performed below 25% accuracy for two of the four categories, and two were below 40% for two of the four categories. None of these extreme cases held for any of the unimpaired patients, whose accuracy rates were higher and were distributed equally across categories. Additionally, when comparing performance between impaired and unimpaired patients across all six learning blocks, these patients only differed in the last two blocks (see Table 2)—a pattern consistent with the non-impaired patients eventually obtaining and maintaining all four categories. By contrast, the same block by impairment analysis for the patients using non-optimal strategies in the II task showed differences between the impaired and unimpaired frontal patients distributed across early and later blocks (Table 2).

With respect to information-integration classification, a framework proposed by Seger (2008) is relevant. Seger suggests that the link between ventral striatum and the medial PFC (orbital PFC

Table 2
Block by block performance for patients adopting non-optimal strategies (impaired) and optimal strategies (non-impaired) for both category structures.

	Mean (SD)					
	Block 1	Block 2	Block 3	Block 4	Block 5	Block 6
II task						
Impaired	0.39 (0.17)	0.43 (0.13)	0.46 (0.14)	0.48 (0.14)	0.54 (0.17)	0.56 (0.14)
Non-impaired	0.45 (0.18)	0.64 (0.11)	0.64 (0.14)	0.69 (0.10)	0.69 (0.15)	0.75 (0.14)
p value	*0.564*	*0.017*	*0.049*	*0.014*	*0.158*	*0.033*
RB task						
Impaired	0.34 (0.11)	0.44 (0.10)	0.49 (0.06)	0.50 (0.05)	0.48 (0.07)	0.47 (0.06)
Non-impaired	0.37 (0.14)	0.52 (0.18)	0.56 (0.19)	0.63 (0.17)	0.70 (0.11)	0.70 (0.11)
p value	*0.73*	*0.448*	*0.478*	*0.128*	*0.006*	*0.001*

and anterior cingulate) forms an important "motivational" loop involved in information-integration category learning. Actor–critic models of reward processing have postulated that the ventral striatum provides a "critical" evaluation of whether the expected reward was received after an action (Joel, Niv, & Ruppin, 2002). With regards to the PFC, functional imaging work has supported the involvement of the medial PFC, in interaction with regions of the striatum, in feedback relative to observation-based learning (Cincotta & Seger, 2007). Furthermore, research demonstrates that patients with damage to medial and orbital PFC have difficulty monitoring feedback and altering strategies in visual association learning (Hornak et al., 2004) and economic (Koenigs & Tranel, 2007) and social (Bechara, Tranel, Damasio, & Damasio, 1996; Moretti, Dragone, & di Pellegrino, 2009) decision-making. In the realm of social decision-making, considerable work in patients with orbital frontal lesions led to the formulation of the somatic marker hypothesis (Bechara, Damasio, & Damasio, 2000), which proposes that regions of orbital PFC are critical for the integration of signals arising from the "body" with the cognitive aspects of decision-making. More recently, Fellows (2007) has argued that many of the deficits seen in vPFC patients with respect to social and economic decision making can be attributed to a deficit in "reversal learning" or the ability to respond to negative feedback by shifting away from sub-optimal, but previously learned, responses. There is some overlap in the framework proposed by Joel et al. (2002) and by Fellows (2007), namely that vPFC is critically involved in the interpretation of feedback signals resulting from actions taken and then motivating changes in the task strategy. In a broader context, multiple sources of converging evidence point to vPFC as critically engaged in integrating internal and external information (in the case of category learning—feedback signals) with goal directed behaviors. In the case of II category learning these signals arise from the ventral striatum and appropriate interpretation of negative feedback would lead to a shift of strategy in order to improve performance. It is this latter ability that we hypothesize may be impaired in our patients with VMPFC damage, who failed to achieve and/or maintain the optimal strategy regardless of category structure after six blocks of performing the task. Clearly, further work will be needed to continue to test this hypothesis.

In addition to the common region of lesion overlap, there were common neuropsychological performance measures in the impaired patients that also did not dissociate the two category learning structures. Patients impaired at II and patients impaired at RB both demonstrated significantly lower levels of performance on the WCST. Given previous formulations of the WCST as a task that requires the explicit use of rules (Eling, Derckx, & Maes, 2008) it is puzzling why the patients' performance on this task did not dissociate across the two category structures, only one of which requires the explicit use of rules. While the ability to generate or maintain rules is part of what is required in the WCST, more recent studies have begun to examine other important components such as response perseveration and learned irrelevance (Maes, Damen, & Eling, 2004; Maes, Vich, & Eling, 2006). In the current case, patients with impaired performance were either unable to adopt, or adopted and could not maintain, the optimal strategy for a given category structure. What makes it particularly difficult for these patients is that a non-optimal strategy often produces a significant amount of positive feedback, so that in the face of such positive reinforcement they would still need to abandon their current approach in order to improve performance. In other words, the current learning strategy, while somewhat effective, nevertheless must be discarded as no longer relevant. The inability to let go of an ineffective strategy would be reflected in response perseverations. Evidence consistent with this notion comes from a closer examination of performance on the WCST. The mean number of categories reached by the impaired group was 3.2, suggesting that these patients were able to master the 3 rules critical to the WCST – color, number and form – but were unable to drop those previously learned categories in the face of negative feedback. Finding the optimal category learning strategy would be dependent on the ability to abandon previous learning in order to achieve a maximal level of performance.

While performance on the WCST showed a relationship with impairment on both II and RB tasks, other neuropsychological measures revealed differences. For instance, only impairment on RB performance showed a significant relationship with VIQ and new verbal learning, as indexed by the Warrington Recognition Memory Test for words. This finding is consistent with formulations of the RB task as involving the explicit learning and implementation of verbal rules (Ashby et al., 1998). Additionally, functional imaging research of RB category learning tasks demonstrates clear involvement of the medial temporal lobe during RB classification (Nomura et al., 2007) and the MTL has been demonstrated to be critical for performing well on tests of explicit memory (Squire, 1992; Moscovitch, Nadel, Winocur, Gilboa, & Rosenbaum, 2006). While none of the frontal patients in this study had additional damage to the MTL, the common lesion location of VMPFC – which has direct connections with MTL through the basal forebrain (Dere, Easton, Nadel, & Huston, 2008) – has been shown to play an important role in episodic recollection (Farovik, Dupont, Arce, & Eichenbaum, 2008) and the ability to make accurate judgments about the accessibility of episodic memories (Schnyer et al., 2004). The latter ability has been postulated to involve a similar mechanism of integrating internally available information in order to guide behavior (Schnyer, Nicholls, & Verfaellie, 2005).

Interestingly, the one patient without VMPFC damage (FL_11) who was impaired on the RB task only, had damage to the left DLPFC and demonstrated the lowest verbal learning ability of all patients. Therefore, the mechanism of impairment of this single patient may have been different than that of the remaining patients who all had VMPFC damage. While only suggestive, the results from the one patient impaired at only the RB task points to the critical role of DLPFC in RB learning, a result consistent with the COVIS model as well as literature pointing to the DLPFC as involved in the implementation of explicit rules (Filoteo et al., 2005a; Muhammad, Wallis, & Miller, 2006). For instance, previous fMRI research implicates regions of DLPFC as responsible for hypothesis testing in category learning (Seger & Cincotta, 2006) and this is consistent with theories that postulate a more general role of DLPFC in rule implementation (Rougier, Noelle, Braver, Cohen, & O', 2005). In addition to patient FL_11, only two other patients in the current study had clear DLPFC lesions (one on the left side and one on the right). A more through examination of the role of DLPFC in RB category learning will require a greater number of patients with damage to this region.

6. Conclusions

The current study examined category learning utilizing two different category structures—one that requires the use of explicit rules and one that involves information-integration, in a group of patients with lesions to prefrontal cortex. Utilizing model-based analysis, a subset of patients was found responsible for the impairment in both RB and II tasks. All had lesions involving regions of VMPFC and were impaired at the WCST. For both category learning structures, the ability to monitor feedback and adjust strategy is critical to learning and this ability is what appears impaired by lesions to VMPFC. However, there were also differences. Those patients impaired at RB learning failed to utilize all categories effectively and this difficulty achieving optimal performance was related to their verbal IQ and new verbal learning ability. By contrast,

patients impaired at II appeared unable to shift and/or maintain their approach to the task and "engage" the procedural learning system that is mediated by the striatum. These results are consistent with a common mechanism that is critical for learning of either category structure and irrespective of whether an explicit, rule-based system or a procedural system is engaged.

Acknowledgements

We would like to thank Ginette Lafleche, Ph.D. and Michael Alexander, M.D. for the neuropsychological testing and assessment of the patients. We would also like to thank Caitlin Tenison, Natalie Dailey, Maria Olivares, and Sasha Wolosin for their assistance in testing normal control participants. This research was supported by NIMH grants MH077708 to WTM, NINDS grant NS047884 to SWE, MH071783 to MV, and Army grant #W911NF-07-2-0023 to DMS through the Center for Strategic and Innovative Technologies at The University of Texas at Austin, and by the Medical Research Service of the Department of Veterans Affairs.

Appendix A.

Three different classes of decision bound models were fit to the final block of data from each participant. In this Appendix, we describe the models that were fit to each participant's responses. We organize this section around the two category structures since every model was not applied to data from every condition.

A.1. Rule-based condition

Hypothesis-testing models. Four models were compatible with the assumption that observers used an explicit hypothesis-testing strategy. The *optimal model* assumes that the observer sets a criterion on the length dimension, sets a criterion on the orientation dimension, and integrates that information post-decisionally. The model assumes that these decision criteria are those that maximize accuracy (i.e., the decision bounds shown in Fig. 1). The optimal model uses the following decision rule: Respond A if the line length is short and the orientation is shallow, Respond B if the line length is short and the orientation is steep, Respond C if the line length is long and the orientation is shallow, Respond D if the line length is long and the orientation is steep. This model has one free parameter: the variance of internal (perceptual and criterial) noise (i.e., σ^2). Three additional hypothesis-testing models that used the same decision rule were tested. The *sub-optimal-length model* assumes that the observer used the optimal decision criterion along the orientation dimension, but used a sub-optimal decision criterion along the length dimension. The *sub-optimal-orientation model* assumes that the observer used the optimal decision criterion along the length dimension, but used a sub-optimal decision criterion along the orientation dimension. These two models contain two free parameters (i.e., one criterion and the noise variance). The *sub-optimal-length-orientation model* assumes that the observer used a sub-optimal decision criterion along the length dimension and a sub-optimal decision criterion along the orientation dimension. This model contains three free parameters (i.e., two decision criteria and the noise variance).

Information-integration models. One information-integration model was fit to the data. The Striatal Pattern Classifier (SPC; Ashby & Waldron, 1999) assumes that there are four "units" in the length-orientation space. On each trial the observer determines which unit is closest to the perceptual effect and gives the associated response. When fitting the SPC to the rule-based condition data, we assume that each category has one associated unit. This model results in four "minimum-distance-based" decision bounds. Because the location of one of the units can be fixed and since a uniform expansion of contraction of the space will not affect the location of the resulting (minimum distance) decision bounds, the model contains six free parameters (i.e., five that determine the location of the units, and one noise variance). This model has been found to provide a good computational model of observers response regions in previous information-integration category learning studies (e.g., Ashby & Waldron, 1999; Ashby, Waldron, Lee, & Berkman, 2001; Maddox, 2001, 2002; Maddox, Ashby, et al., 2004; Maddox, Filoteo, et al., 2004). In addition, the assumptions of this model have strong neurobiological plausibility.

Random-responder models. One model assumes that the participant responds A, B, C, or D with probability 0.25 for each stimulus. This model has no free parameters. A second model estimates the probability of responding A, B, C, and D from the data with the constraint that these probabilities sum to one. This model has three free parameters.

A.2. Information-integration condition

Hypothesis-testing models. Three models were compatible with the assumption that observers used an explicit hypothesis-testing strategy to solve the information-integration category learning problem. The assumptions of the *hypothesis-testing(1) model* are identical to those from the sub-optimal-length-orientation model (described above) and assume that the observer sets a decision criterion along the length dimension, a decision criterion along the orientation dimension, and uses the same post-decisional integration rule outlined above. The *hypothesis-testing(2) model* instantiates an "extreme values" type of decision rule. This model assumes that the observer sets two criteria along the length dimension that partitions the length dimension into short, medium and long line lengths. The model assumes that the observer responds A if the length is short, B if the length is intermediate and the orientation is steep, C if the length intermediate and the orientation is shallow, and D if the length is long. The *hypothesis-testing(3) model* is similar, but it assumes that the observer sets two criteria along the orientation dimension that partitions the orientation dimension into shallow, intermediate, and steep line orientations. The model assumes that the observer responds A if the orientation is intermediate and the length is short, B if the orientation is steep, C if the orientation is shallow, and D if the orientation is intermediate and the length is long. Both of these models contain three free parameters (two criteria and one noise).

Information-integration models. The *optimal model* assumes that the observer used the optimal decision bounds (see Fig. 1) and contains the single noise parameter. The SPC was also applied to the data under the same assumptions used when applying the model to the rule-based condition.

Random-responder models. The same models outlined above were applied.

A.3. Model fits

The relevant models were fit separately to the final block of data for each participant. The model parameters were estimated using maximum likelihood (Ashby, 1992; Wickens, 1982) and the goodness-of-fit statistic was

$$\mathrm{BIC} = r\ \ln(N) - 2\ln L,$$

where r is the number of free parameters, N is the number of trials being fit (100) and L is the likelihood of the model given the data (Akaike, 1974; Takane & Shibayama, 1992). The BIC statistic penalizes a model for extra free parameters in such a way that the smaller the BIC, the closer a model is to the "true model," regard-

less of the number of free parameters. Thus, to find the best model among a given set of competitors, one simply computes a BIC value for each model, and chooses the model associated with the smallest BIC value.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716–723.

Aron, A. R., Shohamy, D., Clark, J., Myers, C., Gluck, M. A., & Poldrack, R. A. (2004). Human midbrain sensitivity to cognitive feedback and uncertainty during classification learning. *Journal of Neurophysiology*, *92*(2), 1144–1152.

Ashby, F. G. (1992). Multivariate probability distributions. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 1–34). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, *105*, 442–481.

Ashby, F. G., & Ell, S. W. (2001). The neurobiology of human category learning. *Trends in Cognitive Science*, *5*(5), 204–210.

Ashby, F. G., & Ennis, J. M. (2006). The role of the basal ganglia in category learning. *The Psychology of Learning and Motivation*, *47*(1–36)

Ashby, F. G., Ennis, J. M., & Spiering, B. J. (2007). A neurobiological theory of automaticity in perceptual categorization. *Psychological Review*, *114*(3), 632–656.

Ashby, F. G., & Gott, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *14*(1), 33–53.

Ashby, F. G., & Maddox, W. T. (1993). Relations between prototype, exemplar, and decision bound models of categorization. *Journal of Mathematical Psychology*, *37*, 372–400.

Ashby, F. G., & Maddox, W. T. (2005). Human category learning. *Annual Reviews in Psychology*, *56*, 149–178.

Ashby, F. G., & Waldron, E. M. (1999). On the nature of implicit categorization. *Psychonomic Bulletin and Review*, *6*(3), 363–378.

Ashby, F. G., Waldron, E. M., Lee, W. W., & Berkman, A. (2001). Suboptimality in human categorization and identification. *Journal of Experimental Psychology: General*, *130*, 77–96.

Bechara, A., Damasio, H., & Damasio, A. R. (2000). Emotion, decision making and the orbitofrontal cortex. *Cerebral Cortex*, *10*(3), 295–307.

Bechara, A., Tranel, D., Damasio, H., & Damasio, A. R. (1996). Failure to respond autonomically to anticipated future outcomes following damage to prefrontal cortex. *Cerebral Cortex*, *6*(2), 215–225.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(433–436).

Brown, J., Bullock, D., & Grossberg, S. (1999). How the basal ganglia use parallel excitatory and inhibitory learning pathways to selectively respond to unexpected rewarding cues. *Journal of Neuroscience*, *19*(23), 10502–10511.

Cincotta, C. M., & Seger, C. A. (2007). Dissociation between striatal regions while learning to categorize via feedback and via observation. *Journal of Cognitive Neuroscience*, *19*(2), 249–265.

Cools, R., Clark, L., Owen, A. M., & Robbins, T. W. (2002). Defining the neural mechanisms of probabilistic reversal learning using event-related functional magnetic resonance imaging. *Journal of Neuroscience*, *22*(11), 4563–4567.

Dere, E., Easton, A., Nadel, L., & Huston, J. P. (2008). *Handbook of episodic memory*. Oxford, UK: Elsevier.

Eling, P., Derckx, K., & Maes, R. (2008). On the historical and conceptual background of the Wisconsin Card Sorting Test. *Brain and Cognition*, *67*(3), 247–253.

Ell, S. W., Marchant, N. L., & Ivry, R. B. (2006). Focal putamen lesions impair learning in rule-based, but not information-integration categorization tasks. *Neuropsychologia*, *44*(10), 1737–1751.

Estes, W. K. (1956). The problem of inference from curves based on group data. *Psychological Bulletin*, *53*, 134–140.

Farovik, A., Dupont, L. M., Arce, M., & Eichenbaum, H. (2008). Medial prefrontal cortex supports recollection, but not familiarity, in the rat. *Jpurnal of Neuroscience*, *28*(50), 13428–13434.

Fellows, L. K. (2007). The role of orbitofrontal cortex in decision making: A component process account. *Ann N Y Acad Sci*, *1121*, 421–430.

Fellows, L. K., & Farah, M. J. (2003). Ventromedial frontal cortex mediates affective shifting in humans: Evidence from a reversal learning paradigm. *Brain*, *126*(Pt 8), 1830–1837.

Filoteo, J.V., Lauritzen, J.S., & Maddox, W.T. (in press). Removing the frontal lobes: The effects of engaging executive functions on perceptual category learning. *Psychological Science*.

Filoteo, J. V., Maddox, W. T., Ing, A. D., Zizak, V., & Song, D. D. (2005). The impact of irrelevant dimensional variation on rule-based category learning in patients with Parkinson's disease. *Journal of the International Neuropsychological Society*, *11*(5), 503–513.

Filoteo, J. V., Maddox, W. T., Simmons, A. N., Ing, A. D., Cagigas, X. E., Matthews, S., et al. (2005). Cortical and subcortical brain regions involved in rule-based category learning. *Neuroreport*, *16*(2), 111–115.

Frank, M. J. (2005). Dynamic dopamine modulation in the basal ganglia: A neurocomputational account of cognitive deficits in medicated and nonmedicated Parkinsonism. *Journal of Cognitive Neuroscience*, *17*(1), 51–72.

Ghods-Sharifi, S., Haluk, D. M., & Floresco, S. B. (2008). Differential effects of inactivation of the orbitofrontal cortex on strategy set-shifting and reversal learning. *Neurobiology of Learning and Memory*, *89*(4), 567–573.

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. *Journal of Neuroscience*, *26*(32), 8368–8376.

Hornak, J., O'Doherty, J., Bramham, J., Rolls, E. T., Morris, R. G., Bullock, P. R., et al. (2004). Reward-related reversal learning after surgical excisions in orbito-frontal or dorsolateral prefrontal cortex in humans. *Journal of Cognitive Neuroscience*, *16*(3), 463–478.

Joel, D., Niv, Y., & Ruppin, E. (2002). Actor-critic models of the basal ganglia: New anatomical and computational perspectives. *Neural Networks: The Official Journal of the International Neural Network Society*, *15*(4–6), 535–547.

Knowlton, B. J., Mangels, J. A., & Squire, L. R. (1996). A neostriatal habit learning system in humans. *Science*, *273*, 245–254.

Koenigs, M., & Tranel, D. (2007). Irrational economic decision-making after ventromedial prefrontal damage: Evidence from the Ultimatum Game. *Journal of Neuroscience*, *27*(4), 951–956.

Lawrence, A. D. (2000). Error correction and the basal ganglia: Similar computations for action, cognition and emotion? *Trends in Cognitive Science*, *4*(10), 365–367.

Lezak, M. D. (1995). *Neuropsychological assessment* (3rd ed.). New York: Oxford University Press.

Lombardi, W. J., Andreason, P. J., Sirocco, K. Y., Rio, D. E., Gross, R. E., Umhau, J. C., et al. (1999). Wisconsin Card Sorting Test performance following head injury: Dorsolateral fronto-striatal circuit activity predicts perseveration. *Journal of Clinical and Experimental Neuropsychology*, *21*(1), 2–16.

Maddox, W. T. (1999). On the dangers of averaging across observers when comparing decision bound models and generalized context models of categorization. *Perception and Psychophysics*, *61*(2), 354–375.

Maddox, W. T. (2001). Separating perceptual processes from decisional processes in identification and categorization. *Perception and Psychophysics*, *63*, 1183–1200.

Maddox, W. T. (2002). Learning and attention in multidimensional identification, and categorization: Separating low-level perceptual processes and high level decisional processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 99–115.

Maddox, W. T., & Ashby, F. G. (1993). Comparing decision bound and exemplar models of categorization. *Perception and Psychophysics*, *53*, 49–70.

Maddox, W. T., & Ashby, F. G. (2004). Dissociating explicit and procedural-learning based systems of perceptual category learning. *Behavioural Processes*, *66*(3), 309–332.

Maddox, W. T., Ashby, F. G., Ing, A. D., & Pickering, A. D. (2004). Disrupting feedback processing interferes with rule-based but not information-integration category learning. *Memory and Cognition*, *32*(4), 582–591.

Maddox, W. T., & Filoteo, J. V. (2005). The neuropsychology of perceptual category learning. In H. Cohen, & C. Lefebvre (Eds.), *Handbook of categorization in cognitive science* (pp. 573–599). Elsevier, Ltd.

Maddox, W. T., & Filoteo, J. V. (2007). Modeling visual attention and category learning in amnesiacs, striatal-damaged patients and normal aging. In *Advances in clinical-cognitive science: Formal modeling and assessment of processes and symptoms*. Washington, DC: American Psychological Association., pp. 113–146.

Maddox, W. T., Filoteo, J. V., Hejl, K. D., & Ing, A. D. (2004). Category number impacts rule-based but not information-integration category learning: Further evidence for dissociable category-learning systems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(1), 227–245.

Maes, J. H., Damen, M. D., & Eling, P. A. (2004). More learned irrelevance than perseveration errors in rule shifting in healthy subjects. *Brain and Cognition*, *54*(3), 201–211.

Maes, J. H., Vich, J., & Eling, P. A. (2006). Learned irrelevance and response perseveration in a total change dimensional shift task. *Brain and Cognition*, *62*(1), 74–79.

Milner, B. (1963). Effects of different brain lesions on card sorting. *Archives of Neurology*, *9*, 90–100.

Monchi, O., Petrides, M., Doyon, J., Postuma, R. B., Worsley, K., & Dagher, A. (2004). Neural bases of set-shifting deficits in Parkinson's disease. *Journal of Neuroscience*, *24*(3), 702–710.

Monchi, O., Petrides, M., Petre, V., Worsley, K., & Dagher, A. (2001). Wisconsin Card Sorting revisited: Distinct neural circuits participating in different stages of the task identified by event-related functional magnetic resonance imaging. *Journal of Neuroscience*, *21*(19), 7733–7741.

Moretti, L., Dragone, D., & di Pellegrino, G. (2009). Reward and social valuation deficits following ventromedial prefrontal damage. *Journal of Cognitive Neuroscience*, *21*(1), 128–140.

Moscovitch, M., Nadel, L., Winocur, G., Gilboa, A., & Rosenbaum, R. S. (2006). The cognitive neuroscience of remote episodic, semantic and spatial memory. *Current Opinions in Neurobiology*, *16*(2), 179–190.

Muhammad, R., Wallis, J. D., & Miller, E. K. (2006). A comparison of abstract rules in the prefrontal cortex, premotor cortex, inferior temporal cortex, and striatum. *Journal of Cognitive Neuroscience*, *18*(6), 974–989.

Nelson, H. E. (1976). A modified card sorting test sensitive to frontal lobe deficits. *Cortex*, (12), 313–324.

Nomura, E. M., Maddox, W. T., Filoteo, J. V., Ing, A. D., Gitelman, D. R., Parrish, T. B., et al. (2007). Neural correlates of rule-based and information-integration visual category learning. *Cerebral Cortex*, *17*(1), 37–43.

Partington, J. E., & Leiter, R. G. (1949). Partington's pathway test. *The Psychological Service Center Bulletin*, *1*, 9–20.
Pelli, D. G. (1997). The Video Toolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442.
Poldrack, R. A., Prabhakaran, S. C. A., & Gabrieli, J. D. (1999). Striatal activation during acquisition of a cognitive skill. *Neuropsychology*, *13*, 564–574.
Price, A. L. (2005). Cortico-striatal contributions to category learning: Dissociating the verbal and implicit systems. *Behavioral Neuroscience*, *119*(6), 1438–1447.
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.
Rougier, N. P., Noelle, D. C., Braver, T. S., Cohen, J. D., & O', R. C. (2005). Prefrontal cortex and flexible cognitive control: Rules without symbols. *Proceedings of the National Academy of Sciences of United States of America*, *102*(20), 7338–7343.
Schnyer, D. M., Nicholls, L., & Verfaellie, M. (2005). The role of VPMC in metamemorial judgments of content retrievability. *Journal of Cognitive Neuroscience*, *17*, 832–846.
Schnyer, D. M., Verfaellie, M., Alexander, M. P., Lafleche1, G., Nicholls, L., & Kaszniak, A. W. (2004). A role for right medial prefrontal cortex in accurate feeling of knowing judgments: Evidence from patients with lesions to frontal cortex. *Neuropsychologia*, *42*(7), 957–966.
Schultz, W. (1998). Predictive reward signal of dopamine neurons. *Journal of Neurophysiology*, *80*(1), 1–27.
Seger, C. A. (2008). How do the basal ganglia contribute to categorization? Their roles in generalization, response selection, and learning via feedback. *Neuroscience and Biobehavioral Reviews*, *32*(2), 265–278.
Seger, C. A., & Cincotta, C. M. (2002). Striatal activity in concept learning. *Cognitive, Affective, and Behavioral Neuroscience*, *2*(2), 149–161.
Seger, C. A., & Cincotta, C. M. (2005). The roles of the caudate nucleus in human classification learning. *Journal of Neuroscience*, *25*(11), 2941–2951.
Seger, C. A., & Cincotta, C. M. (2006). Dynamics of frontal, striatal, and hippocampal systems during rule learning. *Cerebral Cortex*, *16*(11), 1546–1555.
Shaw, M. L. (1982). Attending to multiple sources of information: I. The integration of information in decision making. *Cognitive Psychology*, *14*, 353–409.
Smith, J. D., & Minda, J. P. (1998). Prototypes in the mist: The early epochs of category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 1411–1436.
Squire, L. R. (1992). Memory and the hippocampus: A synthesis from findings with rats, monkeys and humans. *Psychological Review*, *99*, 195–231.
Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*, 643–662.
Stuss, D. T., & Benson, D. F. (1986). *The fronal lobes*. New York: Raven Press.
Takane, Y., & Shibayama, T. (1992). Structures in stimulus identification data. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 335–362). Mahwah, NJ: Erlbaum.
Takahashi, Y. K., Roesch, M. R., Stalnaker, T. A., Haney, R. Z., Calu, D. J., Taylor, A. R., et al. (2009). The orbitofrontal cortex and ventral tegmental area are necessary for learning from unexpected outcomes. *Neuron*, *62*(2), 269–280.
Warrington, E. K. (1984). *Manual for Recognition Memory Test*. Windsor, England: NFER-Nelson.
Wechsler, D. (1997). *Wechsler Adult Intelligence Scale—Third Edition*. San Antonio: Harcourt Brace & Company.
Wickens, T. D. (1982). *Models for behavior: Stochastic processes in psychology*. San Francisco: Freeman.