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 New York University, New York, New York

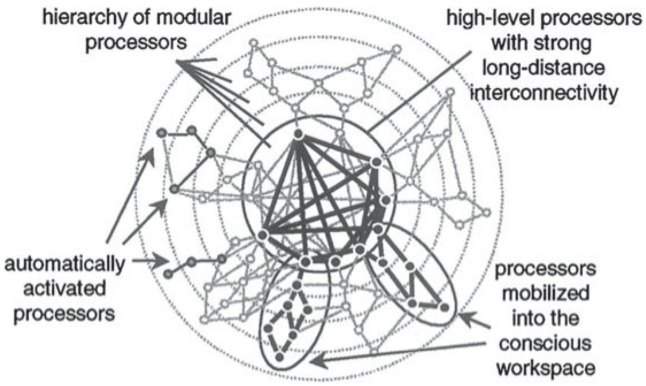

Figure 77.1 Schematic diagram of the global workspace. Sensory activations in the back of the brain are symbolized by dots and lines in the outside rings. Dominant sensory neural coalitions (dark lines and dots) compete with one another to trigger reverberatory activity in the global workspace (located 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 experience 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

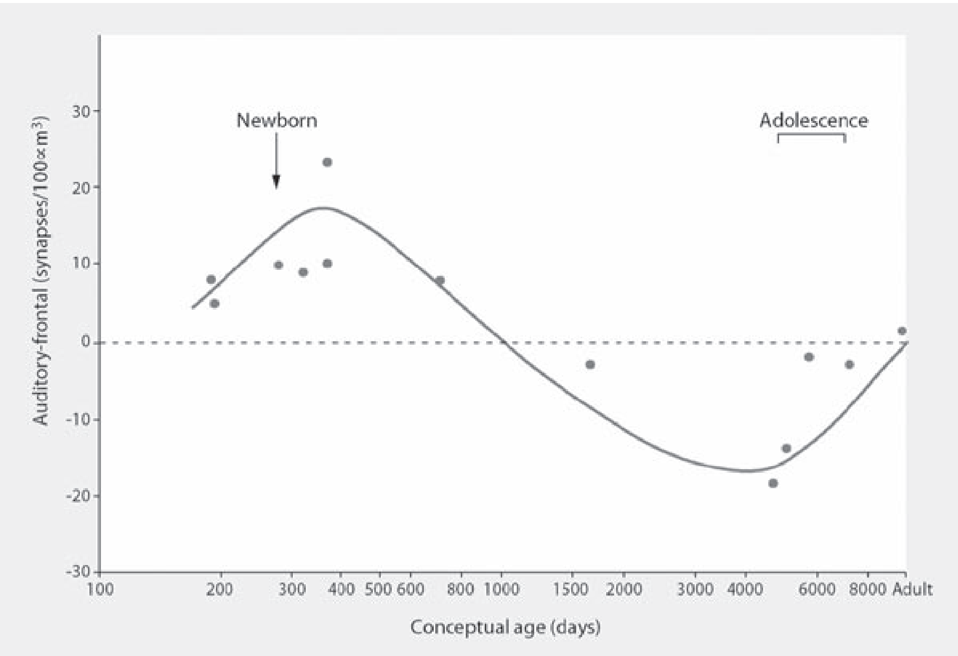

Figure 77.2 Relative synaptic density of auditory and frontal cortex. Conceptual age is age from conception. The peak at the left of roughly three months (postnatal) reflects a high number of auditory synapses relative to frontal synapses. (From Gazzaniga, Ivry, & Mangun, 2002.)

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.