CHAPTER 1

Consciousness

ILYA FARBER AND WILLIAM P. BANKS

Consciousness is an inclusive term for a number of central aspects of our personal existence. It is the arena of self-knowledge, the ground of our individual perspective, the realm of our private thoughts and emotions. It could be argued that these aspects of mental life are more direct and immediate than any perception of the physical world; indeed, it was this view that led Descartes to claim that the fact of our own thinking is the only empirical thing we know with mathematical certainty. Nevertheless, the study of consciousness within science has proven both challenging and controversial, so much so that some have doubted the appropriateness of addressing it within the tradition of scientific psychology. In the last few decades, however, new methods and technologies have yielded striking insights into the nature of consciousness. Neuroscience in particular has begun to reveal detailed connections between brain events, subjective experiences, and cognitive processes. The effect of these advances has been to give consciousness a central role both in integrating the diverse areas of psychology and in relating them to developments in neuroscience. In this chapter, we survey what has been discovered about consciousness; but because of the unique challenges that it poses, we also devote a fair amount of discussion to methodological and theoretical issues and consider ways in which prescientific models of consciousness exert a lingering (and potentially adverse) influence.

Consciousness has two features that pose special methodological challenges for scientific investigation. First, and best known, is its inaccessibility. A conscious experience is directly accessible only to the one person who has it, and even for that person it is often not possible to express precisely and reliably what has been experienced. As an alternative, psychologists have developed indirect measures (such as physiological measurements and reaction time) that permit reliable and quantitative measurement—but at the cost of raising new methodological questions about the relationship between these measures and consciousness itself.

The second challenging feature is that the single word *consciousness* is used to refer to a broad range of related but distinct phenomena (Farber & Churchland, 1995; Hill, 2009). *Consciousness* can mean not being knocked out or asleep; it can mean awareness of a particular stimulus, as opposed to unawareness or implicit processing; it can mean the basic functional state that is modulated by drugs, depression, schizophrenia, or REM sleep. It is the higher-order self-awareness that some species have and others lack; it is the understanding of one's own character and motivations that is gained only after careful reflection; it is the inner voice that expresses some small fraction of

---

Author's Note: Bill Banks passed away before revisions to this chapter could be completed. His coauthor and volume editors dedicate this work to his memory. Bill did much to bring the scientific study of consciousness to its current state of prominence and respectability, including cofounding the journal *Consciousness and Cognition* and organizing the inaugural meeting of the Association for the Scientific Study of Consciousness. He was a cheerful and tireless catalyst for scientific progress, and he will be sorely missed.

what is actually going on below the surface of the mind. On one very old interpretation, it is a transcendent form of unmediated presence in the world; on another, perhaps just as old, it is the inner stage on which ideas and images present themselves in quick succession.

Where scientists are not careful to focus their inquiry or to be explicit about what aspect of consciousness they are studying, this diversity can lead to confusion and talking at cross purposes. On the other hand, careful decomposition of the concept can point the way to a variety of solutions to the *first* problem, the problem of access. As it has turned out, the philosophical problems of remoteness and subjectivity need not always intrude in the study of more specific *forms* of consciousness such as those already mentioned; some of the more prosaic senses of consciousness have turned out to be quite amenable to scientific analysis. Indeed, a few of these, such as "awareness of stimuli" and "ability to remember and report experiences," have become quite central to the domain of psychology, and must now by any measure be considered well studied.

In what follows, we provide a brief history of the early development of scientific approaches to consciousness, followed by more in-depth examinations of the two major strands in 20th-century research: the cognitive and the neuroscientific. In this latter area especially, the pace of progress quickened in the mid-1990s and then accelerated dramatically in the 2000s. Although no single model has yet won broad acceptance, it has become possible for theorists to advance hypotheses with a degree of empirical support and fine-grained explanatory power that was undreamed of 20 years ago (see, e.g., Welshon, 2011). In the concluding section, we offer some thoughts about the relationship between this scientific progress and everyday understanding.

BRIEF HISTORY OF THE STUDY OF CONSCIOUSNESS

Ebbinghaus (1908, p. 3) remarked that psychology has a long past and a short history. The same could be said for the study of consciousness, except that the past is even longer and the scientific history shorter. The concept that the soul is the organ of experience and, hence, the organ of consciousness is ancient. This is a fundamental idea in the Platonic dialogues, as well as in the Upanishads, written about 600 years before Plato wrote and recording thinking that was then already ancient.

We could look at the soul as part of a prescientific explanation of mental events and their place in nature. In the mystical traditions, the soul is conceived as a substance different from the body that inhabits the body, survives its death (typically by traveling to a supernatural realm), and is the seat of thought, sensation, awareness, and usually the personal self. This doctrine is also central to Christian belief, and, for this reason, it has had enormous influence on Western philosophical accounts of mind and consciousness. The doctrine of soul or mind as an immaterial substance separate from body is not universal. Aristotle considered but did not accept the idea that the soul might leave the body and re-enter it (*De Anima*, 406; see Aristotle, 1991). His theory of the different aspects of "soul" is rooted in the functioning of the biological organism. The pre-Socratic philosophers for the most part had a materialistic theory of soul, as did Lucretius and the later materialists, and the conception of an immaterial soul is foreign to the Confucian tradition. The alternative prescientific conceptions of consciousness suggest that many problems of consciousness being faced today are not inevitable consequences of a scientific investigation of awareness. Rather, they may result from the specific assumption that mind and matter are entirely different substances.

The mind–body problem is the legendary and most basic problem posed by consciousness (Dardis, 2008). The question asks how subjective experience can be created by matter, or in more modern terms, by the interaction of neurons in a brain. Descartes (1596–1650; see Descartes, 1951) provided an answer to this question, and his answer formed the modern debate. Descartes's famous solution to the problem is that body and soul are two different substances. Of course, this solution is a version of the religious doctrine that soul is immaterial and has properties entirely different from those of matter. This position is termed dualism, and it assumes that consciousness does not arise from matter at all. The question then becomes not how matter gives rise to mind, because these are two entirely different kinds of substance, but how the two different substances can interact. If dualism is correct, a scientific program to understand how consciousness arises from neural processes is clearly a lost cause, and, indeed, any attempt to reconcile physics with experience is doomed. Even if consciousness is not thought to be an aspect of "soul-stuff," its concept has inherited properties from soul-substance that are not compatible with our concepts of physical causality. These include free will, intentionality, and subjective experience. Further, any theorist who seeks to understand how mind and body "interact" is implicitly assuming dualism. To those who seek a unified view of nature, consciousness under these conceptions creates insoluble problems. The philosopher Schopenhauer called the mind–body problem the “worldknot” because of the seeming impossibility of reconciling the facts of mental life with deterministic physical causality. Chalmers (1996) termed the problem of explaining how brains give rise to subjective experience the “hard problem,” to distinguish it from the supposedly easier problem of explaining particular functions of consciousness.

Gustav Fechner, a physicist and philosopher, attempted to establish (under the assumption of dualism) the relationship between mind and body by measuring mathematical relations between physical magnitudes and subjective experiences of magnitudes (see, e.g., Dzhafarov & Colonius, 2011). Although no one would assert that he solved the mind–body problem, the methodologies he devised to measure sensation helped to establish the science of psychophysics.

The tradition of structuralism in the 19th century, in the hands of Wundt, Titchener, and many others (see Boring, 1942), led to very productive research programs. The structuralist research program could be characterized as an attempt to devise laws for the psychological world that have the power and generality of physical laws, clearly a dualistic project. Nevertheless, many of the “laws” and effects they discovered are still of interest to researchers (see, e.g., Day & Kimm, 2010).

The publication of John Watson’s (1925) book *Behaviorism* marked the end of structuralism in the United States. Methodological and theoretical concerns about the current approaches to psychology had been brewing, but Watson’s critique, essentially a manifesto, was thoroughgoing and seemingly definitive. For some 40 years afterwards, it was widely accepted that psychological research should study only publicly available measures such as accuracy, heart rate, and response time; that subjective or introspective reports were valueless as sources of data; and that consciousness itself could not be studied. Watson’s arguments were consistent with views of science being developed by logical positivism, a school of philosophy that opposed metaphysics and argued that statements were meaningful only if they have empirically verifiable content. They were also consistent with ideas (later expressed by Wittgenstein, 1953, and Ryle, 1949) that people do not have privileged access to the inner workings of our minds through introspection, and, thus, that subjective reports were questionable sources of data. The mind (and the brain) were considered a black box, an area closed to investigation, and all theories were to be based on examination of observable stimuli and responses.

Research conducted on perception and attention during World War II, the development of the digital computer and information theory, and the emergence of linguistics as a scientific study of mind led to changes in every aspect of the field of psychology. It was widely concluded that the behavoristic strictures on psychological research had led to narrow theories of little relevance to interesting aspects of human performance. Chomsky’s blistering attack on behaviorism (reprinted as Chomsky, 1996) might be taken as the 1960s equivalent of Watson’s earlier behavioristic manifesto. Henceforth, researchers in psychology had to face the very complex mental processes demanded by linguistic competence, which were beyond the reach of methods countenanced by behaviorism. The mind was no longer a black box; theories based on a wide variety of techniques were used to develop rather complex theories of what went on in the mind. New theories and new methodologies emerged with dizzying speed, in what was termed the “cognitive revolution” (Gardner, 1985; Proctor & Vu, 2006).

We could consider ourselves to be in a second phase of this revolution, or possibly a new revolution built on the shoulders of the earlier one. This second revolution results from the progress that has been made by techniques that allow researchers to observe processing in the brain, including encephalography (EEG and ERP) and imaging techniques such as positron emission tomography (PET) and functional magnetic resonance imaging (fMRI). This last black box, the brain, is now being opened.

Our best chance of resolving the difficult problems of consciousness, including the “world knot” of the mind–body problem, would seem to come from the increasing depth and breadth of scientific methods available for relating matter (neural structure and function) to mind (psychological measures of perception and cognition). A true solution to the problem of consciousness may await conceptual change, or it may remain always at the tantalizing boundary where science intersects with philosophy; but at the very least, we must count it as progress that we have now entered an era in which the pursuit of questions about awareness, volition, and metacognition can be considered a task of normal science, and can be conducted with increasingly sophisticated technological tools.

WHAT WE HAVE LEARNED FROM MEASURES OF COGNITIVE FUNCTIONING

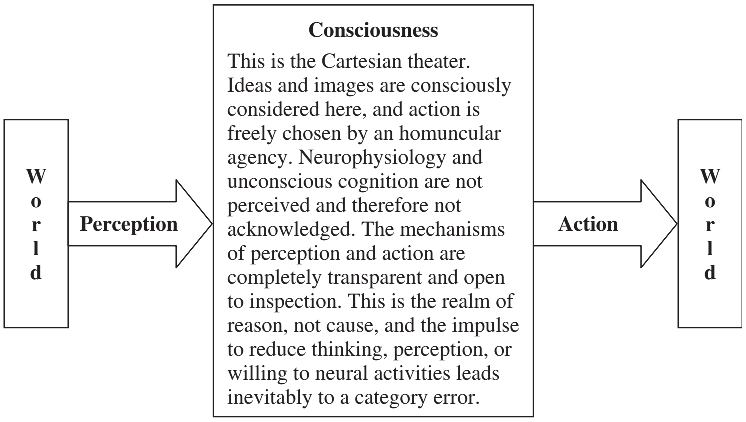
Research on consciousness using strictly behavioral data has a history that long predates the present explosion of knowledge derived from neuroscience. This history includes sometimes-controversial experiments on unconscious or subliminal perception and on influences of consciously unavailable stimuli on performance and judgment. A fresh observer looking over the literature might note wryly that the research is more about unconsciousness than consciousness. Indeed, this is a fair assessment of the research, but it is that way for a good reason.

The motivation for this direction of research can be framed as a test of the folk theory of the role of consciousness in perception and action. A sketch of such a folk theory is presented in Figure 1.1. This model—mind as a container of ideas, with a window to the world for perception at one end and for action at the other—is consistent with a wide range of metaphors about mind, thought, perception, and intention (cf. Lakoff, 1987; Lakoff & Johnson, 1980). The folk model has no room for unconscious thought, and any evidence for unconscious thought would be a challenge to the model. The approach of normal science would be to attempt to disconfirm its assumptions and thus search for unconscious processes in perception, thought, and action.

The folk theory has enormous power because it defines common sense and provides the basis for intuition. In addition, the assumptions are typically implicit and unexamined. For all these reasons, the folk model can be very tenacious. Indeed, as McClosky and colleagues (e.g., McCloskey & Kohl, 1983) showed, it can be very difficult to get free of a folk theory. They found that a large proportion of educated people, including engineering students enrolled in college physics courses, answered questions about physical events by using a folk model closer to Aristotelian physics than Newtonian.

Many intuitive assumptions can be derived from the simple outline in Figure 1.1. For example, the idea (sometimes termed "naïve realism") that perception is essentially a transparent window on the world, unmediated by nonconscious physiological processes, is seen in the direct input from the world to consciousness. The counterpart to naïve realism, which we might call naïve conscious agency, is that actions have, as their sufficient cause, the intentions generated in consciousness, and further, that the intentions arise entirely within consciousness, on the basis of consciously available premises.

The container metaphor is by now so familiar that we can easily forget that it *is* a metaphor. It can take many forms, from a wandering sequence of pictures on a screen to CPU-controlled evolution of instructions stored in RAM. One of the most dominant forms of this metaphor is what Dennett (1991) has dubbed the "Cartesian theater," the idea that consciousness is a sort of inner stage or screen upon which percepts appear. Versions of the container metaphor are deeply entrenched in our everyday language: We say that we *see* an idea (where could it be "viewed" by the mind if not on our inner stage?), we say that we have an idea *in* our mind, we announce that we are putting something *out* of mind, that we are holding an image in our mind's eye, and so on. Perceptions or ideas or intentions are brought forth in the conscious theater, and they are examined and dispatched in "the light of reason." The "machine" model of mental processing (Lakoff & Johnson, 1980) is another common folk model, in which the "thought processing machine" takes the place of the Cartesian stage. The transparency of perception and action is retained, but in that model the process of thought is hidden in the machine and may not be available to consciousness. Both folk models require an observer (homunculus) to supervise operations and make decisions about action.



This is the Cartesian theater. Ideas and images are consciously considered here, and action is freely chosen by an homuncular agency. Neurophysiology and unconscious cognition are not perceived and therefore not acknowledged. The mechanisms of perception and action are completely transparent and open to inspection. This is the realm of reason, not cause, and the impulse to reduce thinking, perception, or willing to neural activities leads inevitably to a category error.

Figure 1.1 A folk model of the role of consciousness in perception and action

As has been pointed out by Churchland (1986, 1996) and Banks (1993), this mental model leads to assumptions that make consciousness an insoluble problem. For example, the connection among ideas in the mind is not causal in this model, but logical, so that the reduction of cognitive processing to causally related biological processes is impossible, philosophically a "category error."[1] Further, the model leads to a distinction between reason (in the mind) and cause (in matter), and, thus, is another route to dualism. The homunculus has free will, which is incompatible with deterministic physical causality. In short, a host of seemingly undeniable intuitions about the biological irreducibility of cognitive processes derive from comparing this model of mind with intuitive models of neurophysiology (which themselves may have unexamined folk-neurological components).

Given that mental processes are, in fact, grounded in neural processes, an important task for cognitive science is to provide a substitute for the model of Figure 1.1 that is compatible with biology. Such a model will likely be as different from the folk model as relativity theory is from Aristotelian physics. Below we consider a number of research projects that are in essence attacks on the model of Figure 1.1.

Unconscious Perception

It goes without saying that a great deal of unconscious processing must take place between registration of stimulus energy on a receptor and perception. This should itself place doubt on the naïve realism of the folk model, which views the entire process as transparent. We do not here consider these processes in general (they are treated in the chapters on sensation and perception) but only studies that have looked for evidence for a possible route from perception to memory or response that does not go through the central theater. We begin with this topic because it raises a number of questions and arguments that apply broadly to studies of unconscious processing.

The first experimentally controlled study of unconscious perception is apparently that of Peirce and Jastrow (1884). They found that differences between lifted weights that were not consciously noticeable were nonetheless discriminated at an above-chance level. Another early study showing perception without awareness is that of Sidis (1898), who found above-chance accuracy in naming letters on cards that were so far away from the observers that they complained that they could see nothing at all. This has been a very active area of investigation. The early research history on unconscious perception was reviewed by Adams (1957). More recent reviews include Dixon (1971, 1981), Bornstein and Pittman (1992), and Baars (1988, 1997), and introductory articles are included in Baars, Banks, and Newman (2003). The critical review of Holender (1986), along with the commentary in the same issue of *Behavioral and Brain Sciences*, contains arguments and evidence that are still of interest.

A methodological issue that plagues this area of research is that of assuring that the stimulus is not consciously perceived. This should be a simple technical matter, but many studies have set exposures at durations brief enough to prevent conscious perception, then neglected to reset them as the threshold lowered over the session because the participants dark-adapted or improved in the task through practice. Experiments that presumed participants were not aware of stimuli out of the focus of attention often did not have internal checks to test whether covert shifts in attention were responsible for perception of the purportedly unconscious material.

Even with perfect control of the stimulus, there is the substantive issue of what constitutes the measure of unconscious perception. One argument would deny unconscious perception by definition: The very finding that performance was above chance demonstrates that the stimuli were not subliminal. The lack of verbal acknowledgment of the stimuli by the participant might come from a withholding of response, from a very strict personal definition of what constitutes "conscious," or have many other interpretations. A behaviorist would have little interest in these subjective reports, and indeed it might be difficult to know what to make of them because they are reports on states observable only to the participant. The important point is that successful discrimination, whatever the subjective report, could be taken as an adequate certification of the suprathreshold nature of the stimuli.

The problem with this approach is that it takes consciousness out of the picture altogether. One way of getting it back in was suggested by Cheesman and Merikle (1984). They proposed a distinction between the objective threshold, which is the point at which performance, by any measure, falls to chance, and the subjective

---

[1] A more scientific-sounding version of this objection is the claim that neuroscientific theories of consciousness show only a "correlation" between neural and mental activity, and, thus, commit the fallacy of mistaking correlation for causation. Despite its intuitive plausibility, this objection has little force against modern theories, and misunderstands both their evidentiary base and their role in the evolution of the concept of consciousness (Farber, 2005).

threshold, which is the point at which participants report they are guessing or otherwise have no knowledge of the stimuli. Unconscious perception would be above-chance performance with stimuli presented at levels falling between these two thresholds. Satisfying this definition amounts to finding a dissociation between consciousness and response. For this reason Kihlstrom, Barnhardt, and Tataryn (1992) suggested that a better term than *unconscious perception* would be *implicit perception,* which has been used by subsequent authors (e.g., Chen & Treisman, 2009; Simons, Hannula, Warren, & Day, 2007) interested in whether a stimulus can affect a response without awareness of the stimulus. This term is used by analogy with implicit memory, which is an influence of memory on performance without conscious recollection of the material itself (see McNamara, this volume). The well-established findings of implicit memory in neurological cases of amnesia make it seem less mysterious that perception could also be implicit.

The distinction between objective and subjective threshold raises a new problem, the measurement of the subjective threshold. Accuracy of response can no longer be the criterion. We are then in the position of asking the person if he or she is aware of the stimulus. Just asking may seem a dubious business, but several authors have remarked that it is odd that we accept the word of brain-damaged people when they claim they are unaware of a stimulus for which implicit memory can be demonstrated, but are more skeptical about the reports of awareness or unawareness by normal participants with presumably intact brains. There is rarely a concern that the participant is untruthful in reporting on awareness of the stimulus. The problem is more basic than honesty; rather, the problem is that awareness is a state that is not directly accessible by the experimenter. A concrete consequence of this inaccessibility is that it is impossible to be sure that the participant shares the experimenter's definition of awareness. Simply asking the participant if he or she is aware of the stimulus amounts to embracing the participant's definition of awareness and probably aspects of the participant's folk model of mind, with all the problems such acceptance of unspoken assumptions entails. It is, therefore, important to find a criterion of awareness that is not subject to experimental biases, assumptions by the participants about the meaning of the instructions, and so on.

Some solutions to this problem are promising. One is to present a stimulus in a degraded form such that the participant reports seeing nothing at all, then test whether some attribute of that stimulus is perceived or otherwise influences behavior. This approach has the virtue of using a simple and easily understood criterion of awareness while testing for a more complex effect of the stimulus. Not seeing anything at all is a very conservative criterion, but it is far less questionable than more specific criteria.

Another approach to the problem has been to look for a qualitative difference between effects of the same stimulus presented above and below the subjective threshold. Such a difference would give converging evidence that the subjective threshold has meaning beyond mere verbal report. In addition, the search for differences between conscious and unconscious processing is itself of considerable interest as a way of assessing the role of consciousness in processing. This is one way of addressing the important question: What is consciousness for? Finding differences between conscious and unconscious processing is a way of answering this question. This amounts to applying the contrastive analysis advocated by Baars (1988; see also James, 1983).

Holender's (1986) criticism of the unconscious-perception literature points out, among other things, that in nearly all of the findings of unconscious perception the response to the stimulus—for example the choice of the heavier weight in the Peirce and Jastrow (1884) study—is the same for both the conscious and the putatively unconscious case. The only difference, then, is the subjective report that the stimulus was not conscious. Because this report is not independently verifiable, the result is on uncertain footing. If the pattern of results is different below the subjective threshold, this criticism has less force.

A dramatic difference between conscious and unconscious influences is seen in the exclusion experiments of Merikle, Joordens, and Stolz (1995). The exclusion technique, devised by Jacoby (1991; cf. Debner & Jacoby, 1994; Jacoby, Lindsay & Toth, 1992; Jacoby, Toth, & Yonilinas, 1993) requires a participant *not* to use some source or type of information in responding. If the information nevertheless influences the response, there seems to be good evidence for a nonconscious effect.

The Merikle et al. (1995) experiment presented individual words, like *spice*, one at a time on a computer screen for brief durations ranging up to 214 milliseconds (ms). After each presentation, participants were shown word stems like *spi-* on the screen. Each time they were asked to complete the stem with any word that had *not* just been presented. Thus, if *spice* was presented, that was the only word they could not use to complete *spi-* (so *spin, spite, spill,* and so on would be acceptable, but not *spice*). They were told that sometimes the presentation would be too brief for them to see anything, but they were asked to do their best. When nothing at all was shown, the stem was completed 14% of the time with one of the prohibited words. This result represents a baseline percentage. The proportion at 29 ms was 13.3%, essentially the baseline level. This performance indicates that 29 ms is below the objective threshold because it was too brief for there to be any effect at all, and, of course, also below the subjective threshold, which is higher than the objective threshold.

The important finding is that, with the longer presentations of 43 and 57 ms, there was an *increase* in use of the word that was to be excluded. Finally it returned *below* baseline to 8% at 214 ms. The interpretation of this result is that at 43 and 57 ms the word fell above the objective threshold, so that it was registered at some level by the nervous system and associatively primed *spice.* However, at these durations it was below the subjective threshold so that its registration was not conscious, and it could not be excluded. Finally, at the still longer duration of 214 ms, it was frequently above the subjective threshold, and it could be excluded.

This set of findings suggests a hypothesis about the function of consciousness that is applicable in many domains, namely, that with consciousness of a stimulus comes the ability to control how it is used. This could only be discovered in cases in which there was some registration of the stimulus below the subjective threshold, as was the case here.

The only concern with this experiment is that the subjective threshold was not independently measured. To make the argument complete, there should be a parallel measurement of the subjective threshold. It would be necessary to show independently that the threshold for conscious report is between 57 and 214 ms. This particular criticism does not apply to some similar experiments, such as those by Cheesman and Merikle (1986).

Finally, whatever the definition of consciousness, or of the subjective threshold, there is the possibility that the presented material was consciously perceived, if only for an instant, and then, the fact that it had been conscious was forgotten. If this were the case, the folk model in which conscious processing is necessary for any cognitive activity to take place is not challenged. It is very difficult to test the hypothesis that there was a brief moment of forgotten conscious processing that did the cognitive work being attributed to unconscious processing. It may be that this hypothesis is untestable, but testable or not, it seems implausible as a general principle. Complex cognitive acts like participating in a conversation and recalling memories take place without awareness of the cognitive processing that underlies them. If brief moments of immediately forgotten consciousness were nonetheless the motive power for all cognitive processing, it would be necessary to assume that everyone is afflicted with a dense amnesia, conveniently affecting only certain aspects of mental life. It seems more parsimonious to assume that these mental events were never conscious in the first place.

Perceptual Construction

We tend to think of an object of perception—the thing we are looking at or hearing—as an entity with coherence, a single representation in the mind. However, this very coherence has become a theoretical puzzle because the brain does not represent an object as a single entity. Rather, specialized areas of the brain analyze various aspects of the object separately, and a single object or image is nowhere to be found. How the brain keeps parts of an object together is termed the binding problem, as will be discussed later in the section on neurophysiology. Here we cover some aspects of the phenomenal object and what it tells us about consciousness.

Rock (1983) presented a case for there being a "logic of perception," a system of principles by which perceptual objects are constructed. The principles are themselves like "tools" and are not available to awareness. We can only infer them by observing the effects of appropriate displays on perception. One principle we learn from ambiguous figures is that we can see only one interpretation at a time. There exist many bi-stable figures, such that one interpretation is seen, then the other, but never both. Logothetis and colleagues have shown that both interpretations are represented simultaneously in the brain, competing in such a way that only one at a time can yield a conscious percept (Logothetis & Sheinberg, 1996).

Rock (1983) suggested that unconscious assumptions determine which version of an ambiguous figure is seen, and by extension, he would argue that this is a normal component of perception of unambiguous objects. Real objects seen under normal viewing conditions typically have only one interpretation, and there is no way to show the effect of interpretation so obvious with ambiguous figures. Because the "logic of perception" is not conscious, the folk model of naïve realism does not detect a challenge in this process; all that one is aware of is the result, and its character is attributed to the object rather than to any unconscious process that may be involved in its representation (see Klatzky & Lederman, this volume, and Proffitt & Caudek, this volume).

The "New Look" in perceptual psychology (Erdelyi, 1974; McGinnies, 1949) attempted to show that events that are not registered consciously, as well as unconscious expectations and needs, can influence perceptions or even block them, as in the case of perceptual defense. Bruner (1992) pointed out that the thrust of the research was to demonstrate higher-level cognitive effects in perception, not to establish that there were nonconscious ones. However, unacknowledged constructive or defensive processes would necessarily be nonconscious.

The thoroughgoing critiques of the early New Look research program (Eriksen, 1958, 1960, 1962; Fuhrer & Eriksen, 1960; Neisser, 1967) cast many of its conclusions in doubt, but they had the salubrious effect of forcing subsequent researchers to avoid many of the methodological problems of the earlier research. Better controlled research by Shevrin and colleagues (Bunce, Bernat, Wong, & Shevrin, 1999; Shevrin, 2000; Wong, Bernat, Bunce, & Shevrin, 1997) suggests that briefly presented words that trigger defensive reactions (established in independent tests) are registered but that the perception is delayed, in accord with the older definition of perceptual defense.

One of the theoretical criticisms (Eriksen, 1958) of perceptual defense was that it required a "superdiscriminating unconscious" that could prevent frightening or forbidden images from being passed on to consciousness. Perceptual defense was considered highly implausible because it would be absurd to have two complete sets of perceptual apparati, especially if the function of one of them were only to protect the other from emotional distress. If a faster unconscious facility existed, so goes the argument, there would have been evolutionary pressure to have it be the single organ of perception and thus of awareness. The problem with this argument is that it assumes the folk model summarized in Figure 1.1, in which consciousness is essential for perception to be accomplished. If consciousness were not needed for all acts of perception in the first place, then it is possible for material to be processed fully without awareness, acted upon in some manner, and only selectively to become available to consciousness.

Bruner (1992) suggested as an alternative to the superdiscriminating unconscious the idea of a *Judas Eye*, which is a term for the peephole a speakeasy bouncer uses to screen out the police and other undesirables. The Judas Eye would be a process that uses a feature to filter perception, just as the speakeasy peephole can permit discrimination based on only a glimpse of a uniform or a badge. However, there is evidence that unconscious detection can rely on relatively deep analysis. For example, Mack and Rock (1998) found that words presented without warning while participants were judging line lengths (a difficult task) were rarely seen. This is one of several phenomena they termed "inattentional blindness." On the other hand, when the participant's name or a word with strong emotional content was presented, it was reported much more frequently than neutral words. (Detection of one's name from an unattended auditory source has been reported in much-cited research; see Cowan & Wood, 1997; Lamy et al., this volume; Wood & Cowan, 1995a, 1995b.) Because words like *rape* were seen and visually similar words like *rope* were not, the superficial visual analysis of a Judas Eye does not seem adequate to explain perceptual defense and related phenomena. It seems a better hypothesis that there is much parallel processing in the nervous system, most of it unconscious, and some products become conscious only after fairly deep analysis.

Another "object" to consider is the result of memory construction. In the model of Figure 1.1, the dominant metaphor for memory is recalling an object that is stored in memory. It is as though one goes to a "place" in memory where the "object" is stored, and then brings it into consciousness. William James referred to the "object" retrieved in such a manner as being "as fictitious . . . as the Jack of Spades." There is an abundance of modern research supporting James. Conscious memory is best viewed as a construction based on pieces of stored information, general knowledge, opinion, expectation, and so on (one excellent source to consult on this is Schacter, 1995). Neisser (1967) likened the process of recall to the work of a paleontologist who constructs a dinosaur from fragments of fossilized bone, using knowledge derived from other reconstructions. The construction aspect of the metaphor is apt, but in memory, as in perception, we do not have a good model of what the object being constructed is or what the neural correlate is. The folk concept of a mental object, whether in perception or in memory, may not have much relation to what is happening in the nervous system when something is perceived or remembered.

Subliminal Priming and Negative Priming

Current interest in subliminal priming derives from Marcel's work (1983a, 1983b). His research was based on earlier work showing that perception of one word can "prime" a related word (Meyer & Schvaneveldt, 1971; see McNamara, this volume). The primed word is processed more quickly or accurately than in control conditions without priming.

Marcel (1983a, 1983b) reported a series of experiments in which he obtained robust priming effects in the absence of perception of the prime. His conclusion was that priming, and, therefore, perception of the prime word proceeds automatically and associatively, without any necessity for awareness. The conscious model (cf. Figure 1.1) would be that the prime is consciously registered, serves as a retrieval cue for items like the probe, and, thus, speeds processing for probe items. Marcel presented a model in which consciousness serves more as a monitor of psychological activity than as a critical path between perception and action. Holender (1986) and others have criticized this work on a number of methodological grounds, but subsequent research has addressed most of his criticisms (see Kihlstrom et al., 1992, for a discussion and review of this work).

Other evidence for subliminal priming includes Greenwald, Klinger, and Schuh's (1995) finding that the magnitude of affective priming does not approach zero as detection of the priming word approaches zero (see also Draine & Greenwald, 1998). Shevrin and colleagues demonstrated classical conditioning of the galvanic skin response (GSR) to faces presented under conditions that prevented detection of the faces (Bunce et al., 1999; Wong et al., 1997).

Cheesman and Merikle (1986) reported a dissociation of conscious and unconscious priming effects using a variation of the Stroop (1935) interference effect. In the Stroop effect a color word such as *red* is printed in a color different from the one named, in this case, for example, blue. When presented with this stimulus, *red* printed in blue, the participant must say, "blue." Interference is measured as a much longer time to pronounce "blue" than if the word did not name a conflicting color.

Cheesman and Merikle (1986) used a version of the Stroop task in which a word printed in black is presented briefly on a computer screen, then removed and replaced with a colored rectangle the participant is to name. Naming of the color of the rectangle was slowed if the color word named a different color. They then showed, first, that if the color word was presented so briefly that the participant reported having seen nothing, naming of the color was still slowed. This would be classified as a case of unconscious perception, but because the same direction of effect is found both consciously and unconsciously, there would be no real dissociation between conscious and unconscious processing. Holender (1986) and other critics could argue reasonably that it was only shown that the Stroop effect was fairly robust at very brief durations, and the supplementary report of awareness by the participant is unrelated to processing.

Cheesman and Merikle (1986) devised a clever way to answer this criticism. The procedure was to arrange the pairs such that the word *red* would be followed most of the time by a blue rectangle, the word *blue* would be followed by a yellow rectangle, and so on. This created a predictive relationship between the word and the color that participants could strategically exploit to make the task easier. They apparently did use these relationships in naming the colors. With clearly supraliminal presentation of the word, a reversal in the Stroop effect was found such that the red rectangle was named faster when preceded by the word *blue* than when preceded by the word *red*.

However, this reversal was found only when the words were presented for longer than the duration needed to perceive them. When the same participants saw the same sequence of stimuli with words that were presented too briefly for conscious perception, they showed only the normal Stroop effect. The implication of this result is that the sort of interference found in the Stroop effect is an automatic process that does not require conscious perception of the word. What consciousness of the stimulus adds is control. Only when there was conscious registration of the stimulus could the participants use the stimulus information strategically.

Negative priming is an interference, measured in reaction time or accuracy, in processing a stimulus that was previously presented but was irrelevant to the task decision. It was first discovered by Dalrymple-Alford and Budayr (1966) in the context of the Stroop effect. They showed that if a participant was given, say, the word *red* printed in blue, an immediately following instance of red would take longer to name than other colors.

Negative priming has been found in many experiments in which the negative prime, while presented supraliminally, is not consciously perceived because it is not attended. Tipper (1985) presented overlapping line drawings, one drawn in red and the other in green. Participants were told to name only the item in one of the colors and not the other. After participants had processed one of the drawings, the one they had excluded was sometimes presented on the next trial. In these cases the previously unattended drawing was slower to name than in a control condition in which it had not been previously presented. Banks, Roberts, and Ciranni (1995) presented pairs of words simultaneously to the left and to the right ears. Participants were instructed to repeat aloud only the word presented to one of the ears. If a word that had been presented to the unattended ear was presented in the next pair to be repeated, the response was delayed.

As mentioned in the previous section, material perceptually available but not attended is often the subject of “inattentional blindness,” that is, it seems to be excluded from awareness. The finding of negative priming suggests that ignored material is perceptually processed and represented in the nervous system, but is evidenced only by its negative consequences for later perception, not by any record that is consciously available.

A *caveat* regarding the implication of the negative priming findings for consciousness is that a number of researchers have found negative priming for fully attended stimuli (MacDonald & Joordens, 2000; Milliken, Joordens, Merikle, & Seiffert, 1998). These findings imply that negative priming cannot be used as evidence by itself that the perception of an item took place without awareness.

Priming studies have been used to address the question of whether the unconscious is, to put it bluntly, “smart” or “dumb.” This is a fundamental question about the role of consciousness in processing; if unconscious cognition is dumb, the function of consciousness is to provide intelligence when needed. If the unconscious is smart—capable of doing a lot on its own—it is necessary to find different roles for consciousness.

Greenwald (1992) argued that the unconscious is dumb because it could not combine pairs of words in his subliminal priming studies. He found that some pairs of consciously presented words primed other words on the basis of a meaning that could only be gotten by combining them. For example, presented together consciously, words like *key* and *board* primed *computer*. When presented for durations too brief for awareness they primed *lock* and *wood,* but not *computer*. On the other hand, Shevrin and Luborsky (1961) found that presenting subliminally pictures of a pen and a knee resulted in subsequent free associations that had *penny* represented far above chance levels. The resolution of this difference may be methodological, but there are other indications that unconscious processing may in some ways be fairly smart even if unconscious perception is sometimes a bit obtuse. Kihlstrom (1987) reviews many other examples of relative smart unconscious processing.

A number of subliminal priming effects have lingered at the edge of experimental psychology for perhaps no better reason than that they make hard-headed experimentalists uncomfortable. One of these is subliminal psychodynamic activation (Silverman, 1983). Silverman and others (see Weinberger, 1992, for a review) have found that subliminal presentation of the single sentence “Mommy and I are one” has a number of objectively measurable positive emotional effects (when compared to controls such as “People are walking” or “Mommy is gone”). It is a frequent criticism that the studies did not make sure that the stimulus was presented unconsciously. However, many psychologists would be surprised if the effects were found even with clearly consciously perceived stimuli. It is possible, in fact, that the effects depend on unconscious processing, and it would be interesting to see if the effects were different when subliminal and clearly supraliminal stimuli are compared.

Implicit Memory

Neurological cases brought this topic to the forefront of memory research, with findings of preserved memory in amnesics (Schacter, 1987). The preserved memory is termed implicit because it is a tacit sort of memory, memory that is discovered in use, not memory that is consciously retrieved or observed. Amnesic patients would, for example, work each day at a Tower of Hanoi puzzle, and each day assert that they had never seen it before, but still show improvement in speed of completing it (Cohen, Eichenbaum, Deacedo, & Corkin, 1985). The stem completion task of Merikle et al. (1995) is another type of implicit task. After the word *spice* was presented, even though the word was not consciously registered, its probability of use would be increased. In a memory experiment amnesics and normals who could not recall the word *spice* would nevertheless be more likely to use it to complete the stem than if it had not been presented.

Investigation of implicit memory in normals quickly led to an explosion of research, which is covered in McNamara, this volume; Marsh & Roediger, this volume; and Johnson, this volume.

Nonconscious Basis of Conscious Content

Earlier, we discussed how the perceptual object is a product of complex sensory processes and probably of inferential processes as well. Memory has also been shown to be a highly inferential skill, and the material retrieved from memory is as much a product of inference as it is of retrieval. These results violate an assumption of the folk model by which objects are not constructed but are simply brought into the central arena, whether from perception or memory. Errors of commission in memory serve much the same evidentiary function in memory as ambiguous figures in perception, except they are much more common and easier to induce. The sorts of error we make in eyewitness testimony, or as a result of a number of documented memory illusions (Loftus, 1993), are particularly troublesome because they are made—and believed—with certainty. Legitimacy is granted a memory on the basis of a memory's clarity, completeness, quantity of details, and other internal properties, and the possibility that it is the result of suggestion, association, or other processes is considered invalidated by the internal properties. (Henkel, Franklin, & Johnson, 2000). Completely bogus memories, induced by an experimenter, can be believed with tenacity (cf. Schacter, 1995; see also Marsh & Roediger, this volume, and Roediger & McDermott, 1995).

The Poetzl phenomenon is the reappearance of unconsciously presented material in dreams, often transformed so that the dream reports must be searched for evidence of relation to the material. It has been extended to reappearance in free associations, fantasies, and other forms of output, and a number of studies appear to have found Poetzl effects with appropriate controls and methodology (Erdelyi, 1972; Ionescu & Erdelyi, 1992). Still, the fact that reports must be interpreted and base rates for certain topics or words are difficult to assess casts persistent doubt over the results, as do concerns about experimenter expectations, the need for double-blind procedures in all studies, and other methodological issues.

Consciousness, Will, and Action

In the folk model of consciousness (see Figure 1.1) a major inconsistency with any scientific analysis is the free will or autonomous willing of the homunculus. The average person will report that he or she has free will, and it is often a sign of mental disorder when a person complains that his or her action is constrained or controlled externally. The problem of will is as much of a hard problem (Chalmers, 1996) as is conscious experience. How can willing be put in a natural-science framework?

One approach comes from measurements of the timing of willing in the brain. Libet (Libet, 1985, 1993; Libet et al., 1964; Libet, Alberts, & Wright, 1967) found that changes in EEG potentials recorded from the frontal cortex began 200 to 500 ms before the participant was aware of deciding to begin an action (flexion of the wrist) that was to be done freely. One interpretation of this result is that the perception we have of freely willing is simply an illusion, because by these measurements it comes after the brain has already begun the action.

There are other interpretations that do not lead to this conclusion. The intention that ends with the motion of the hand must have its basis in neurological processes, and it is not surprising that the early stages are not present in consciousness. Consciousness has a role in willing because the intention to move can be arrested before the action takes place (Libet, 1993), and because participation in the entire experimental performance is a conscious act. The process of willing would seem to be an interplay between executive processes, memory, and monitoring, some of which we are conscious and some not. Only the dualistic model of a completely autonomous will controlling the process from the top, like the Cartesian soul fingering the pineal gland from outside material reality, is rejected. Having said this, we must go on to say that a great deal of theoretical work is needed in this area.

The idea of unconscious motivation dates to Freud and before. "Freudian slips" (Freud, 1965), in which unconscious or suppressed thoughts intrude on speech in the form of action errors, should constitute a challenge to the simple folk model by which action is transparently the consequence of intention. However, the commonplace cliché that one has made a "Freudian slip" seems to be more of a verbal habit than a recognition of unconscious determinants of thought, because unconscious motivation is not generally recognized in other areas.

Wegner (1994) has studied some paradoxical (but embarrassingly familiar) effects that result from attempting to suppress ideas. In what they term "ironic thought suppression," they find that the suppressed thought can pose a problem for control of action. Participants trying to suppress a word were likely to blurt it out when speeded in a word association task. Exciting thoughts (about sex) could be suppressed with effort, but they tended to burst into awareness later. The irony of trying to suppress a thought is that the attempt at suppression primes it, and then more control is needed to keep it hidden than if it had not been suppressed in the first place. The intrusion of unconscious ideation in a modified version of the Stroop task (see Baldwin, 2001) indicates that the suppressed thoughts can be completely unconscious and still have an effect on processing.

Attentional Selection

In selective attention paradigms, the participant is instructed to attend to one source of information and not to others that are presented. For example, in the shadowing paradigm, the participant hears one verbal message with one ear and a completely different one with the other. *Shadowing* means "to repeat verbatim a message one hears," and that is what the participants do with one of the two messages. This paradigm has been extremely fruitful, helping attention research to become one of the most important areas in cognitive psychology.

People generally have little awareness of the message on the ear not shadowed (Cherry, 1957; Cowan & Wood, 1997). What happens to that message? Is it lost completely or is some processing performed on it unconsciously? Results showed that participants responded to their name on the unattended channel (Moray, 1959) and would switch the source they were shadowing if the material switched source (Treisman, 1964). Both of these results suggest that unattended material is processed to at least some extent. In the visual modality, Mack and Rock (1998) reported that, in the "inattentional blindness" paradigm, a word presented unexpectedly when a visual discrimination is being conducted is noticed infrequently, but if that word spells the participant's name or an emotional word, it is noticed much more often. For there to be a discrimination between one word and another on the basis of their meaning and not any superficial feature such as length or initial letter, the meaning must have been extracted.

Theories of attention differ on the degree to which unattended material is processed. Early selection theories assume that the rejected material is stopped prior to identification (cf. Broadbent, 1958), with the consequence that unattended material could only be monitored by switching or time sharing. Late selection theories (Deutsch & Deutsch, 1963) assume that unattended material is processed to some depth, perhaps completely, but limitations of capacity prevent it from being perceived consciously or remembered. The results of the processing are available for a brief period and can serve to summon attention or bias the interpretation of attended stimuli, or have other effects. One of these effects would be negative priming, as discussed earlier. Another effect would be the noticing of one's own name or an emotionally charged word from an unattended source.

One set of experiments is consistent with the late selection model, although there are alternative explanations. In an experiment that required somewhat intrepid participants, Corteen and Wood (1972) associated electric shocks with words to produce a conditioned GSR. After the conditioned response was established, the participants performed a shadowing task in which the shock-associated words were presented to the unattended ear. The conditioned response was still obtained, and galvanic skin responses were also obtained for words semantically related to the conditioned words. This latter finding is particularly interesting because the analysis of the words would have to go deeper than just the sound to elicit these associative responses. Other reports of analysis of unattended material include those of Corteen and Dunn (1974), Forster and Govier (1978), MacKay (1973), and Von Wright, Anderson, and Stenman (1975). On the other hand, Wardlaw and Kroll (1976), in a careful series of experiments, did not replicate the effect.

Replication of this effect may be less of an issue than the concern over whether it implies unconscious processing. This is one situation in which momentary conscious processing of the nontarget material is not implausible. There are several lines of evidence that support momentary conscious processing. For example, Dawson and Schell (1982), in a replication of Corteen and Wood's experiment, found that if participants were asked to name the conditioned word in the nonselected ear, they were sometimes able to do so. This suggests that there was attentional switching, or at least some awareness, of material on the unshadowed channel. Corteen (1986) agreed that this was possible. Treisman and Geffen (1967) found that there were momentary lapses in shadowing of the primary message when specified targets were detected in the secondary one. MacKay's (1973) results were replicated by Newstead and Dennis (1979) only if single words were presented on the unshadowed channel and not if words were embedded in sentences. This finding suggests that occasional single words could attract attention and give rise to the effect, while the continuous stream of words in sentences did not create the effect because they were easier to ignore.

Dissociation Accounts of Some Unusual and Abnormal Conditions

The majority of psychological disorders, if not all, have important implications for consciousness, unconscious processing, and so on. Here we consider only disorders that are primarily disorders of consciousness, that is, dissociations and other conditions that affect the quality or the continuity of consciousness, or the information available to consciousness.

Problems in self-monitoring or in integrating one's mental life about a single personal self occur in a variety of disorders. Frith (1992) has described many of the symptoms schizophrenics exhibit as a failure in attributing their actions to their own intentions or agency. In illusions of control, for example, a patient may assert that an outside force made him do something like strip off his clothes in public. By Frith's account this assertion would result from the patient being unaware that he had willed the action, in other words, from a dissociation between the executive function and self-monitoring. The source of motivation is attributed to an outside force ("the Devil made me do it"), when it is only outside the self system of the individual. For another example, schizophrenics are often found to be subvocalizing the very voices that they hear as hallucinations (Frith, 1992, 1996); hearing recordings of the vocalizations does not cause them to abandon the illusion. There are many ways in which the monitoring could fail (see Proust, 2000), but the result is that the self system does not "own" the action, to use Kihlstrom's (1992, 1997) felicitous term.

This lack of ownership could be as simple as being unable to remember that one willed the action, but that seems too simple to cover all cases. Frith's theory is sophisticated and more general. He hypothesized that the self system and the source of willing are separate neural functions that are normally closely connected. When an action is willed, motor processes execute the willed action directly, and a parallel process (similar to feed forward in control of eye movements; see Festinger & Easton, 1974) informs the self system about the action. In certain dissociative states, the self system is not informed. Then, when the action is observed, it comes as a surprise, requiring explanation. Alien hand syndrome (Chan & Liu, 1999; Inzelberg, Nisipeanu, Blumen, & Carasso, 2000) is a radical dissociation of this sort, often connected with neurologic damage consistent with a disconnection between motor planning and monitoring in the brain. (See Proctor & Vu, this volume; and Heuer & Massen, this volume.) In this syndrome, the patient's hand will sometimes perform complex actions, such as unbuttoning the patient's shirt, while he or she watches in horror.

Classic dissociative disorders include fugue states, in which at the extreme the affected person will leave home and begin a new life with amnesia for the previous one, often after some sort of trauma (this may happen more often in Hollywood movies than real life, but it does happen). In all these cases, the self is isolated from autobiographical memory (see Marsh & Roediger, this volume). Dissociative identity disorder is also known as multiple personality disorder. There has been doubt about the reality of this disorder, but there is evidence that some of the multiple selves do not share explicit knowledge with the others (Nissen, Ross, Willingham, Mackenzie, & Schacter, 1994), although implicit memories acquired by one personality seem to be available to the others.

Hysterical dissociations, such as blindness or paralysis (now termed conversion disorders) are very common in wartime or other civil disturbance. One example is the case of 200 Cambodian refugees found to have psychogenic blindness (Cooke, 1991). It was speculated that the specific form of the conversion disorder they had was a result of seeing terrible things before they escaped from Cambodia. Whatever the reason, the disorder could be described as a blocking of access of the self system to visual information, that is, a dissociation between the self and perception. One piece of evidence for this interpretation is the finding that a patient with hysterical analgesia in one arm reported no sensations when stimulated with strong electrical shocks but did have normal changes in physiological indices as they were administered (Kihlstrom et al., 1992). Thus, the pain messages were transmitted through the nervous system and had many of the normal effects, but the conscious monitoring system did not "own" them and so they were not consciously felt.

Anosognosia (Galin, 1992; Ramachandran, 1995, 1996) is a denial of deficits after neurological injury. This denial can take the form of a rigid delusion, defended with tenacity and resourcefulness. Ramachandran et al. (1996) report the case of Mrs. R., a right-hemisphere stroke patient who denied the paralysis of her left arm. Ramachandran asked her to point to him with her right hand, and she did. When asked to point with her paralyzed left hand, the hand remained immobile, but she insisted that she was following the instruction. When challenged, she said, "I have severe arthritis in my shoulder, you know that, doctor. It hurts."

Bisiach and Geminiani (1991) report the case of a woman suddenly stricken with paralysis of the left side who complained on the way to the hospital that a left hand had been forgotten by another patient and left on the ambulance bed. She was able to agree that the left shoulder was hers, and the upper arm, but she became evasive about the forearm and continued to deny the hand altogether.

Denials of this sort are consistent with a dissociation between the representation of the body part or the function (Anton's syndrome is denial of loss of vision, for example) and the representation of the self. Because anosognosia is specific to the neurological conditions (almost always right-hemisphere damage), it is difficult to argue that the denial comes from an unwillingness to admit the deficit. Anosognosia is rarely found with equally severe paralysis resulting from left-hemisphere strokes. (See the section Observations from Human Pathology, further on, for more on the neurological basis for anosognosia and related dissociations.)

Vaudeville and circus sideshows are legendary venues for extreme and ludicrous effects of hypnotic suggestion, such as blisters caused by pencils hypnotically transformed to red-hot pokers, or otherwise respectable people clucking like chickens and protecting eggs they thought they laid on stage. It is tempting to reject these performances as faked, but extreme sensory modifications can be replicated under controlled conditions (Hilgard, 1977). The extreme pain of cold-pressor stimulation can be completely blocked by hypnotic suggestion in well-controlled experimental situations. Recall of a short list of words learned under hypnosis can also be blocked completely by posthypnotic suggestion. In one experiment, Kihlstrom (1994) found that large monetary rewards were ineffective in inducing recall, much to the bewilderment of the participants, who recalled the items quite easily when suggestion was released but the reward was no longer available.

Despite several dissenting voices (Barber, 2000), hypnotism does seem to be a real phenomenon of extraordinary and verifiable modifications of consciousness. Hilgard's (1992) neodissociation theory treats hypnosis as a form of dissociation whereby the self system can be functionally disconnected from other sources of information, or even divided internally into a reporting self and a "hidden observer."

One concern with the dissociative or any other theory of hypnosis is the explanation of the power of the hypnotist. What is the mechanism by which the hypnotist gains such control over susceptible individuals? Without a good explanation of the mechanism of hypnotic control, the theory is incomplete, and any results are open to dismissive speculation. We suggest that the mechanism may lie in a receptivity to control by others that is part of our nature as social animals. By this account, hypnotic techniques are shortcuts to manipulating, for a brief time but with great force, the social levers and strings that are engaged by leaders, demagogues, peers, and groups in many situations.

What Is Consciousness For?

Baars (1988, 1997) suggests that a contrastive analysis is a powerful way to discover the function of consciousness. If unconscious perception does take place, what are the differences between perception with and without consciousness? We can ask the same question about memory with and without awareness. To put it another way, what does consciousness add? As Searle (1992, 1993) points out, consciousness is an important aspect of our mental life, and it stands to reason that it must have some function. What is it?

A few regularities emerge when the research on consciousness is considered. One is that strategic control over action and the use of information seems to come with awareness. Thus, in the experiments of Cheesman and Merikle (1986) or Merikle et al. (1995), the material presented below the conscious threshold was primed but could not be excluded from response as well as it could when presentation was above the subjective threshold. As Shiffrin and Schneider (1977) showed, when enough practice is given to make detection of a given target automatic (that is, unconscious), the system becomes locked into that target and requires relearning if target identity is changed. Automaticity and unconscious processing preserve capacity when they are appropriate, but the cost is inflexibility. These results also suggest that consciousness is a limited-capacity medium, and the choice in processing is between, on the one hand, awareness, control, and limited capacity or, on the other hand, automaticity, unconsciousness, and large capacity.

Another generalization is that consciousness and the self are intimately related. Dissociation from the self can lead to unconsciousness, and conversely, unconscious registration of material can cause it not to be "owned" by the self. This is well illustrated in the comparison between implicit and explicit memory. Implicit memory performance is automatic and not accompanied by a feeling of the sense that "I did it." Thus, after seeing a list containing the word *motorboat* the amnesic completely forgets the list or even the fact that he saw a list, but when asked to write a word starting with *mo*__, he uses *motorboat* rather than more common responses such as *mother* or *moth*. When asked why he used *motorboat*, he would say "I don't know. It just popped into my mind." The person with normal memory who supplies a stem completion that was primed by a word no longer recallable would say the same thing: "It just popped into my head." The more radical lack of ownership in anosognosias is a striking example of the disconnection between the self and perceptual stimulation. Hypnosis may be a method of creating similar dissociations in unimpaired people, so that they cannot control their actions, or find memory recall for certain words blocked, or not feel pain when electrically shocked, all because of an induced separation between the self system and action or sensation.

We could say that consciousness is needed to bring material into the self system so that it is owned and put under strategic control. Conversely it might be said that consciousness emerges when the self is involved with cognition. In the latter case, consciousness is not *for* anything, but rather reflects the fact that what we call conscious experience is the product of engagement of the self with cognitive processing which could otherwise proceed unconsciously. This leaves us with the difficult question of defining the self.

Conclusions

An important advance in the study of consciousness would be to replace the model of Figure 1.1 with something more compatible with findings on the function of consciousness. There are several examples to consider. Schacter (1987) proposed a parallel system with a conscious monitoring function. Marcel's (1983a, 1983b) proposed model is similar in that the conscious processor is a monitoring system. Baars's (1988, 1997) "Global Workspace Theory" seems to be the most completely developed model of this type (see also Franklin & Graesser, 1999, for an artificial intelligence model based on this approach), with parallel processors doing much of the cognitive work, and a self system that has executive functions. We will not attempt a revision of Figure 1.1 more in accord with the current state of knowledge, but any such revision would have parallel processes, some of which are and some of which are not accessible to consciousness. The function of consciousness in such a picture would be controlling processes, monitoring activities, and coordinating the activities of disparate processors. Such an intuitive model might be a better starting point, but we are far from having a rigorous, widely accepted model of consciousness.

Despite the continuing philosophical and theoretical difficulties in defining the role of consciousness in cognitive processing, the study of consciousness may be the one area that offers some hope of integrating the diverse field of cognitive psychology. Virtually every topic in the study of cognition, from perception to motor control, has an important connection with the study of consciousness. Developing a unified theory of consciousness could be a mechanism for expressing how these different functions could be integrated. In the next section, we examine the impact of the revolution in neuroscience on the study of consciousness and cognitive functioning.

NEUROSCIENTIFIC APPROACHES TO CONSCIOUSNESS

The evolution of neuroscientific methodologies for studying consciousness has been strongly influenced by broader technology-driven trends in neuroscience. Prior to the late 20th century, the two most fertile sources of evidence were scalp-based measures of electrical potential (EEG and ERP) and studies of localized brain damage in human patients. EEG, which provides fine temporal resolution but only very weak information about the anatomical *source* of the measured activity, lent itself well to approaches that treated the brain as a single entity, and focused on changes in its functional state over time. The most well known of these is the distinction among the various stages of sleep, but electroencephalography has also yielded insights into finer-grained distinctions within the waking state, such as those between alert and idle or meditative states. The lesion-based approach lent itself more naturally to the search for localizations of function, but while this led to dramatic successes in the areas of sensation, language, and motor control, results for consciousness were far more limited, yielding only relatively crude or ambiguous findings: most notably, that splitting the connections between the hemispheres of the brain seemed to split consciousness (Gazzaniga, 2005), that damage to various midbrain structures can cause total loss of consciousness, and that damage to (or disconnection via lobotomy of) prefrontal cortex can cause surprisingly subtle impairments of conscious self-awareness.

In the 1980s and 1990s, advances in implantable electrode technology led to a new emphasis on recording and theorizing about neural activity at the level of individual cells. Because opportunities to record from implanted electrodes in human subjects are extremely rare, this led in turn to increased efforts to find animal models for studying at least some aspects of consciousness, and work on primates in particular became increasingly important. At the level of theory, this period saw the birth of approaches that sought to identify consciousness in trajectories of activity in neural pathways or in the firing patterns of distributed networks of neurons.

From the late 1990s through the present, the other major trend has been the increasing availability and sophistication of techniques for noninvasive neuroimaging, most notably functional magnetic resonance imaging (fMRI) and magnetoencephalography (MEG). At the same time there has been dramatic progress in noninvasive stimulation via transcranial magnetic stimulation (TMS). Together these methods have rejuvenated the search for anatomical localizations, making it possible to identify consciousness with transient changes in activity in particular areas. The prevalence of neuroimaging has also exerted powerful indirect effects on the broader debate about consciousness. The visually compelling images and (relatively) simple logic of fMRI localization make it a very media-friendly experimental methodology, and as a result the 21st-century public has been exposed to an unprecedented flood of reports about the relationships between brain areas and cognitive functions. Though we are aware of no formal studies of the effect of such exposure on people's beliefs about consciousness, at an anecdotal level it seems that the general public is now far more receptive than before to the idea that consciousness is a function of the brain. Likewise, both in lay thought and in academic philosophy, objections to neuro-materialistic explanations of consciousness as somehow threatening or deflationary have gone from dominating the discourse about consciousness in the 1980s to being little more than a footnote to most contemporary discussions. We see this as a salutary development, because it has enabled students of consciousness to turn their focus to more specific and illuminating questions about the mechanisms, dynamics, and functions of consciousness.

Single-Cell Recording

One of the most compelling lines of research grew out of Nikos Logothetis's discovery that there are single cells in macaque visual cortex whose activity is well-correlated with the monkey's conscious perception (Logothetis, 1998; Logothetis & Schall, 1989). Logothetis' experiments were a variant on the venerable feature-detection paradigm. Traditional feature-detection experiments involve presenting various visual stimuli to a monkey while recording, via an implanted electrode, the activity of a single cell in some particular area of visual cortex. Much of what is known about the functional organization of visual cortex was discovered through such studies; to determine whether a given area is sensitive to, say, color or motion, experimenters vary the relevant parameter while recording from single cells, and look for cells that show consistently greater response to a particular stimulus type.

However, the fact that a single cell represents some visual feature does not necessarily imply anything about what the animal actually perceives; some features extracted by early visual areas (such as center-surround patches) have no direct correlate in conscious perception, and much of the visual system can remain quite responsive to stimuli in an animal anesthetized into unconsciousness. The contribution of Logothetis and his colleagues was to explore the distinction between what is *represented* by the brain and what is *perceived* by the organism. They did so by presenting monkeys with "rivalrous" stimuli, that is, stimuli that support multiple, conflicting interpretations of the visual scene. One common rivalrous stimulus involves two fields of lines flowing past each other; humans exposed to this stimulus report that the lines fuse into a grating that is either upward-moving or downward-moving, and that the perceived direction of motion tends to reverse approximately once per second.

In area MT, which is known to represent visual motion, some cells will respond continuously to a particular stimulus pattern (e.g., an upward-moving grating) for as long as it is present. Within this population of neurons, a subpopulation was found that showed a fluctuating response to rivalrous stimuli, and it was shown that the activity of these cells was correlated with the monkey's behavioral response. For example, within the population of cells that responded strongly to upward-moving gratings, there was a subpopulation whose activity fluctuated (approximately once per second) in response to a rivalrous grating, and whose periods of high activity were correlated with the monkey's behavioral reports of seeing an upward-moving grating.

This discovery was something of a watershed, in that it established that the activity of sensory neurons is not always explicable solely in terms of distal stimulus properties. Comparing the trials in which a given neuron is highly active with those in which it is less active, no difference can be found in the external stimulus or the experimental condition. The only difference that tracks the activity of the cell is the monkey's behavioral *report* about its perception of motion. One might propose that the cells are somehow tracking the monkey's motor output or intention, but given their location and connectivity, this would be hard to support. The most natural interpretation is that these neurons reflect, and perhaps form the neural basis for, the monkey's *awareness* of visual motion.

Some single-cell research seems to show a direct effect of higher-level processes, perhaps related to awareness or intentionality, on lower-level processes. For example, Moran and Desimone (1985) showed that a visual cell's response is modified by the monkey's attentional allocation in the cell's receptive field.

Observations from Human Pathology

One major drawback of single-cell studies is that they are performed almost exclusively on nonhuman animals, because the procedure is invasive and there is little clinical use for single-cell data from a patient's visual cortex. Recent advances in neuroimaging (most notably the advent of fMRI) have made it possible to observe the normal human brain noninvasively, at a fine scale, and in realtime; work of this sort will be discussed later.

Traditionally, however, most of what we know about the functional architecture of the human brain has come from the study of patients who have suffered brain damage, whether from a stroke, an injury, degenerative disease, or surgical excision of cancerous tissue or epileptic foci. Data about the effects of a lesion is gathered from clinical observation and behavioral tests, and this is then correlated with an anatomical characterization of the lesion via MRI or autopsy.

It is famously difficult to use lesion data to ground claims about the localization of function, because a lesion in a given area may disrupt a function without the area itself being in any sense *for* that function—as when the lesion interrupts a pathway, or interferes with other processes necessary for control or behavioral expression. In the case of disruptions related to consciousness, however, merely coming to understand the character of the deficit itself can provide insight into the functional structure of consciousness; just seeing what sorts of breakdowns are *possible* in a system can reveal much about its architecture. Perhaps the clearest example of this has been the phenomenon of "blindsight."

Blindsight occurs in some patients who have suffered damage to primary visual cortex (also known as striate cortex, or area V1 in humans). This damage produces a blind field in the patient's vision, on the side opposite to the lesion; patients will report a complete absence of visual perception in this field. Nonetheless, some patients show a preserved ability to respond in certain ways to stimuli in this field. For example, patients may be able to press a button when a stimulus appears, or to point reliably in the direction of the stimulus, or even to respond appropriately to the emotional content of facial expressions (de Gelder, Vroomen, Pourtois, & Weiskrantz, 1999), all while insisting that they cannot see anything and are "just guessing." Research in humans and monkeys (Weiskrantz, 1990, 1998) has supported the hypothesis that this preserved discriminatory capacity is due to "extrastriate" pathways that bypass primary visual cortex and carry some visual information to other areas of the brain, including areas involved in sensorimotor coordination.

Blindsight relates to the study of consciousness in a number of ways. First, it provides a powerful reminder of how much work goes on outside consciousness; even a form of sensory processing that *results in* a conscious reaction (such as the emotional response to a facial expression, or the diffuse sense that "something has changed") may be quite independent of the sensory information available to consciousness. Second, it clearly demonstrates a functional division, seen throughout the motor system, between the mechanisms involved in consciously selecting and initiating an action and the unconscious mechanisms that guide its implementation and execution (Llinás, 2001). And third, it offers the tantalizing possibility—just beginning to be realized—of using neuroimaging to investigate the differences in activity when the same task is performed with or without conscious awareness (Morris, DeGelder, Weiskrantz, & Dolan, 2001).

Another fruitful line of investigation has involved a constellation of puzzling deficits associated with unilateral damage to parietal cortex. Parietal cortex plays an essential role in coordinating action with perception, and is known to contain a variety of sensory and motor maps that are integrated in complex ways. Right parietal lesions produce partial or complete paralysis of the left side of the body, and almost always produce some degree of *hemineglect*, a tendency to ignore the side of the world opposite the lesion (i.e., the left side; hemineglect is not associated with left parietal lesions). The disorder has both sensory and motor components: Patients will fail to respond to stimuli coming from objects located on the left, and will not spontaneously use their left-side limbs. This lateral bias tends to manifest itself across a variety of modalities and coordinate frames (auditory and visual, body centered and object centered). Many of the standard tests of hemineglect are based on paper-and-pencil tasks carried out with the right hand: For example, patients with the disorder who are asked to copy a picture (presented entirely in the patient's right field) will fill in the right half but leave the left half sketchy or blank, and if asked to bisect a horizontal line they will show a substantial rightward bias. (For a review of clinical and experimental findings regarding hemineglect, see Kerkhoff, 2001).

A variety of mechanisms had been proposed for hemineglect, but the field was narrowed considerably by an ingenious experiment performed by Edoardo Bisiach and his colleagues (Bisiach & Luzzatti, 1978). To discern whether the deficit was primarily one of sense perception or of higher-level processes such as attention and representation, Bisiach designed a test that required only verbal input and output. He asked his subjects to imagine that they were standing at the north end of a well-known plaza in their city, and to recount from memory all the buildings that faced on the plaza. What he found was that patients displayed hemineglect even for this *imagined* vista; in their descriptions, they accurately listed the buildings on the west side of the plaza (to their imaginary right), and omitted some or all of the buildings to the east. Even more strikingly, when he then asked them to repeat the same task but this time to imagine themselves at the south end of the plaza, the left-neglect persisted, meaning that they listed the buildings they had previously omitted and failed to list the same buildings that they had just described only moments before. Because the subjects were drawing on memories formed before the lesion occurred, Bisiach reasoned that the pattern of deficit could only be explained by a failure at the representational level.

This alone would be fascinating, but what makes hemineglect particularly relevant for the study of consciousness is its frequent association with more bizarre derangements of bodily self-conception. For example, some hemineglect patients suffer from misoplegia, a failure to acknowledge that the limbs on the left side of their body are their own. Misoplegic patients often express hatred of the foreign limbs and wish to be rid of them; Ramachandran reports the case of a patient who kept falling out of bed in his attempts to escape his own arm, which he thought was a cadaver's arm placed in his bed by prankish medical students (Ramachandran & Blakeslee, 1998). Other patients, while regarding the limb with indifference, will make bizarre and nonsensical claims such as that it "belongs to" someone else even though it is attached to their own body. It is important to emphasize that these patients are not otherwise cognitively impaired; their IQs are undiminished, and they test at or near normal on tasks that do not involve using or reasoning about the impaired hemifield.

An even stranger disorder associated with hemineglect is anosognosia,[2] or "unawareness of deficit." Anosognosic patients exhibit a near-total unawareness of their paralysis. Though confined to a wheelchair, they will insist that they are capable of full normal use of their left limbs; if pressed, they may produce confabulatory excuses about being "tired" or, in one striking case, "[not] very ambidextrous" (Ramachandran, 1995). Ramachandran has shown that this unawareness even extends to unconscious decisions such as how to grasp or manipulate an object: Anosognosic subjects will use their one good hand to approach tray-lifting or shoe-tying in a way that cannot succeed without help from the other hand, and will either fail to register their failure at the task or will be surprised by it. Bisiach has shown that anosognosia also extends to the perceptual realm; unlike patients with hemifield blindness due to retinal or occipital damage, anosognosics will insist that they are fully functional even when they are demonstrably incapable of responding to stimuli in half of their visual field (Bisiach & Rusconi, 1990).

Anosognosia is a fascinating and puzzling deficit, to which no brief summary will do justice. For our purposes, however, three features are most salient. First and most important is its cognitive impenetrability: Even very intelligent and cooperative patients cannot be made to understand the nature of their deficit. This qualifies the disorder as a derangement of consciousness, because it concerns the subject's inability to form even an *abstract representation* of a particular state of affairs. Second is the bizarre, possibly hallucinatory degree of confabulation associated with the disorder. These confabulations raise deep questions about the relationships between self-perception, self-understanding, and self-description. Third, it should be noted that anosognosia is often strongly domain specific; patients unaware of their paralysis may still admit to other health problems, and double dissociations have been demonstrated between anosognosias for different forms of neglect in single patients (e.g., sensory versus motor neglect, or neglect for personal versus extrapersonal space).

There are at least three major hypotheses about the mechanism of hemineglect and its associated disorders: Bisiach treats it as a systematic warping or "metric distortion" in the patient's representational space (Bisiach, Cornacchia, Sterzi, & Vallar, 1984); Heilman and Schacter attribute it to the failure of second-order monitoring systems (Heilman, Barrett, & Adair, 1998; Schacter, 1990); and Ramachandran presents a complex theory in which the left hemisphere is specialized for building coherence and the right hemisphere (damaged in these disorders) is specialized for using conflicting data to overthrow old interpretations (Ramachandran, 1995). Ramachandran's theory is highly speculative, but is nevertheless the only one to account directly for the stranger cognitive failures of misoplegia and anosognosia. The other theories are not incompatible with the phenomena, but to provide a satisfactory explanation of patients' behavior they would (at minimum) need to be integrated with an account of the mechanisms of confabulation (see, e.g., Moscovitch & Melo, 1997). In any case, what we want to emphasize here is the way in which a lesion of a somatosensory area can produce domain-specific failures of *rationality*. This suggests two counterintuitive ideas about abstract reasoning, a function that has long been assumed to be a function of the frontal lobes: Either it is in fact more broadly distributed across other areas of the brain, including the temporal and parietal cortices or coherent second-order reasoning *about* some domain may require the intact functioning

---

[2] This name is sometimes used in a broader sense, to include the unawareness of other deficits such as amnesia or jargon aphasia. For present purposes we will focus on anosognosia for hemineglect and hemiparesis, since it remains unclear to what extent the broader range of cases can or should be explained in a unitary fashion.

of the areas that construct first-order representations *of* that domain. This second hypothesis would accord well with many recent models of the neural basis of consciousness, in particular those of Damasio and Edelman (discussed later).

The Search for a Theory

Several factors have supported the current flowering of neuroscientific research into consciousness. Tremendous advances in neuroimaging have produced new insight into the functional anatomy of the brain; studies of the response properties of neurons, both *in vitro* and *in vivo* as well as via computer models, have led to a deeper understanding of the principles of neurodynamics. This more sophisticated understanding of the brain has made possible more specific hypotheses about the structures that give rise to consciousness. The search for a neural theory of consciousness can also be seen as part of a broader push for large-scale theories to explain such fundamental brain functions as representation, sensorimotor integration, and executive control (Koch & Davis, 1994). These projects are ambitious, to be sure, but at this point there can be no doubting their scientific respectability.

In the following sections we will consider a number of recent hypotheses. There has been a striking convergence among the major theories of the neural basis of consciousness, a convergence both in conceptual structure and in the choice of brain structures on which to focus. As a consequence, rather than treating individual *theories* one by one, each section below will be devoted to a particular *concept* or theoretical *component* that may play a role in several different theories. There is a trade-off here, because in focusing on the fundamental concepts we must necessarily gloss over some of the details of individual views. We made this choice with an eye to offsetting the limitations of other available treatments of this material: Many of the theorists covered here have published lucid, book-length expositions of their individual views, but we have seen almost no extended synthetic treatments. It is our hope that the approach pursued here will assist the reader both in understanding the individual views and in assessing their contributions to the overall pursuit of consciousness.

Two points about all these theories are worth noting in advance. First, their convergence affords some grounds for optimism that the broad outline of a stable, mature theory of consciousness may be coming into view. The specific current theories of consciousness are doubtless flawed in many respects; but it seems increasingly clear that they are at least looking in the right *place*, and this is a very important step in the development of a scientific subdiscipline. In a nutshell, it seems that the neuroscience of consciousness is on the cusp of moving from a revolutionary to an evolutionary mode of progress. (We will return to this point in the Conclusion to this chapter.)

Second, it is worth briefly noting that all these theories necessarily assume that consciousness is not epiphenomenal—in other words, they treat consciousness as something that plays a functional role and (presumably) confers some concrete advantage on the organisms that have it. This assumption has historically been controversial, but the very success of these theories in guiding empirical research has made epiphenomenalist objections harder to sustain.

Dynamic Activity Clusters

Arguably the first scientific approach to consciousness was that of associationist psychology, which treated consciousness metaphorically as a container or space in which various ideas came and went. The two basic questions posed by the associationists remain with us today: How are ideas formed, and what principles guide the transition from one idea to another? Posing these questions within the context of neuroscience opens, for the first time, the possibility of going beyond the surface level to ask about the *mechanisms* underlying the formation and transition of ideas. Theorists of consciousness are now in a position to ask *how* and even *why* conscious experience is generated, rather than just describing *what* happens in experience.

Many influential theories in this area have shared the basic idea that individual percepts and concepts have as their neural correlate a dynamic "cluster" or "assembly" of neurons (Crick & Koch, 1995; Edelman & Tononi, 2000; Greenfield, 1995; Llinás, Ribary, Contreras, & Pedroarena, 1998; Singer, 1996). Unfortunately there is no broadly accepted blanket term for this set of ideas, so for present purposes we will refer to them as cluster theories. Cluster theories take as their starting point the challenge of distinguishing conscious mental activity from unconscious neural processing. In the sensory systems in particular, it is clear that the brain represents far more information than a person is conscious of at any given moment; for example, the entire visual field is represented in visual cortex, but conscious experience is (usually, more or less) restricted to one small part of that field. What, then, is the neural marker of this distinction? What determines *which* neural representations become, so to speak, the contents of consciousness?

Cluster theories propose that various potentially conscious percepts and/or ideas compete to enter consciousness. Each cluster is a group of neurons, often distributed across multiple areas, which collectively represent some image or sensation. As the brain processes inputs and also recursively processes its own state, different clusters may become active, and some sort of winner-take-all competition determines which one will be most active and (therefore) the object of consciousness. A crucial feature of this hypothesis is that clusters are dynamic and distributed—meaning that a single cluster may incorporate related feature-representations from many different areas of cortex, and a given neuron may participate in different clusters at different times.

Some of the central dynamics of cluster theories are inherited directly from classical associationism, and gain plausibility from the associationist tendencies of neural networks. For example, it is a natural feature of most neural representations that activation will spread from some elements in a cluster to the others, so that activating some features of a representation will cause the network to fill in the missing features, eventually activating the whole cluster. Conversely, the most fundamental principle of learning at the neural level, the idea that neurons that are active at the same time become more strongly connected ("neurons that fire together wire together"), provides a mechanism for the *creation* of clusters on the basis of long-term regularities in experience.

In inheriting this much of the structure of associationism, however, cluster theories also inherit many of its classical problems. It's difficult to give more than a handwaving explanation of how the various contributions of the senses, memory, and imagination interact; the mechanism of intentional conscious direction of thought is likewise obscure. Perhaps most important for the present generation of theories, there are also problems that arise when one tries to characterize the difference between conscious and unconscious representation. Greenfield (1995) explains the difference in terms of magnitude of activation—one is conscious of whichever cluster is most active at a given time—but this is problematic, because magnitude (in the form of firing rate) is already used by the brain to represent the intensity of stimuli. This is reminiscent of the problem critics raised with Locke's claim that memories were distinguished from perception by their faintness; if true, this would mean that the memory of a bright object should be subjectively indistinguishable from the perception of a sufficiently dim object, and this is clearly not the case. If a system is to incorporate *both* a representation of the objective magnitude of a stimulus *and* a distinction between conscious and unconscious representations, that system will need separate ways of encoding these two things; a single variable such as firing rate can't do the job by itself. In the following sections, we will mention some concrete proposals for what additional variables the brain might use for this purpose.

Sensory Imagery and the Binding Problem

At the neural level, one way of interpreting consciousness is as an integration or "binding" of disparate neural representations into a single, coherent percept. When we see an object, its various features such as color, shape, location, movement, and identity are represented in different areas of the brain, but our experience is still *of* a single, unified object that combines all these properties. How is this combination achieved, and how do we know which features go with which object? Christoph von der Malsburg coined the term "binding problem" (von der Malsburg, 1981) to refer to this puzzle in the context of models of the visual system, and it has since been broadened to refer to cross-modal and sensorimotor integration and even to the integration of perception with memory.

As von der Malsburg has pointed out, one can, in principle, solve this problem by having the processing chain terminate in a set of object-specific neurons that stand for whole percepts.[3] This type of representation is highly inefficient and fragile, however, and unsurprisingly the brain does not appear to be organized this way; there is no Cartesian Theater (Dennett, 1991), no single region on which all inputs converge to produce one master representation. Recasting the binding problem, then, the challenge is to explain how a person can have a single, integrated experience of an object whose various properties are represented in different brain regions and *never* brought together in one place.

If not one place, how about one time? An interesting hypothesis that gained prominence in the 1990s is that temporal *synchrony* is what binds representations across the brain (Joliot, Ribary, & Llinás, 1994; Singer, 1996, 2001). The idea here is that all the neurons representing a given percept will produce spikes that closely coincide.

---

[3] This is the type of representation often caricatured as involving "grandmother cells," because, at its most extreme, it would require that there be a single cell for each recognizable object of perception (e.g., your grandmother), and that that cell would fire when and only when you detect that object with any of your senses.

This approach exploits the fact that spike *frequency* does not exhaust the information-carrying potential of a neuronal spike train. Even if two neurons produce the same number of spikes within a given time interval, there are several important ways in which their spike trains may differ. Synchrony thus offers one way to encode the extra representational dimension that cluster theories need. There are also a number of good theoretical reasons to look in this direction, including the following (modified from Singer, 1996):

- The constraints of real-time perceptual processing are such that the mechanism of binding has to work on a very short timescale. It also has to allow for the dynamic creation of novel perceptual clusters involving elements that have never been associated before. Both these requirements suggest that binding should be implemented at the level of neuronal activity rather than at the level of anatomical structure and connectivity.
- It is a robust general principle of neural dynamics that two neurons stimulating a third will have a greater total effect if both their pulses reach the target at the same time (within some small window of tolerance). From this it follows that synchronous firing would enhance the neural "visibility" and associative power of the disparate components of a cluster. Binding via synchrony could thus explain why a visual field containing the same set of features will call up different associations depending on how the features are grouped into objects.
- Neurons in many areas throughout the brain can exhibit oscillatory firing patterns. Phase-locking such oscillations—coordinating them so that their peaks coincide—would be a powerful and efficient means of generating synchrony across large distances in cortex. The need for a mechanism of synchrony would thus provide one (though by no means the only) possible explanation for the ubiquitousness of these oscillatory firing patterns.

The details of empirical studies on synchrony are beyond the scope of this work, but it is now widely accepted that synchronous oscillation plays an important role in visual binding, and may also be crucial for attentional processes and working memory (see Engel, Fries, Konig, Brecht, & Singer, 1999, and Engel & Singer, 2001, for review and discussion). Synchrony can thus be considered at least a neural *precondition* for consciousness, because conscious attention and awareness operate within the realm of whole, bound objects (Treisman & Kanwisher, 1998).

Koch and Crick (1994) advanced a more specific proposal, which has come to be known as the "40-Hz hypothesis." The central idea of this proposal was that synchronous oscillation in the gamma-band frequency range (approximately 25–55 Hz) is both necessary and sufficient for consciousness: In other words, we are conscious of the representational contents of all neurons synchronously oscillating in this frequency band, and all of our conscious imagery is accompanied by such oscillations.

The 40-Hz hypothesis was a breakthrough in two respects. First, it led directly to clear, empirically testable claims about the neural correlate of consciousness (NCC). At the single-cell level, the hypothesis implies that the activity of a given sensory neuron should match the contents of sensory consciousness (e.g., in experiments like those of Logothetis, see earlier) whenever it is oscillating at frequencies in the gamma band. At the level of functional areas, it also follows that consciousness should be insensitive to differences in activity that are restricted to areas that do not exhibit significant gamma-band oscillation. This latter idea was the basis for the famous conjecture that we are not conscious of the contents of V1, the first stage of processing in visual cortex (Crick & Koch, 1995). Unfortunately, as Crick and Koch have themselves pointed out, there are complicating factors that render these seemingly simple implications problematic: How can one distinguish the neurons that are *driving* an oscillation from those that are merely responding to it? What about local inhibitory neurons, which play no direct role in communicating with other cortical areas: Should they be considered part of the NCC if they oscillate? In recognition of these complexities, Crick and Koch shifted to an approach that assumes that the anatomical side of the NCC story will be more complex, involving (at minimum) finer-grained analysis of the contributions of different cell types and cortical layers (Crick & Koch, 1998, 2003), and, more recently, Koch has described synchronized oscillations as just one of several possible candidates for the NCC (Koch, 2004; Tononi & Koch, 2008)

The original 40-Hz hypothesis was novel in a second way that has been less widely noticed, but that may ultimately have more lasting consequences: Unlike previous synchrony models of binding, it provided a way to draw a distinction *within* the realm of bound representations, between those that are and are not conscious. If the 40-Hz hypothesis were correct, a neuroscientist observing the activity of a pair of sensory neurons in separate areas could place them into one of three categories based solely on the properties of their spike trains:

1. If oscillating synchronously in the gamma band, the neurons must be contributing to a single conscious representation.
2. If oscillating synchronously at a frequency *outside* the gamma band, they must be part of a bound representation that is *not* present to consciousness (e.g., an object in an unattended part of the visual field).
3. If active but not oscillating or oscillating out of synchrony, they must be representing features that are unbound, or perhaps bound to different representations.

Even if it now seems that gamma synchrony alone will not be able to fill this role, it would clearly be useful to have *some* means of drawing this distinction between bound representations that are and are not conscious. Another candidate for this function will be discussed in the next section.

Thalamocortical Loops

The thalamus is a lower forebrain structure that is sometimes referred to as the "gateway to the brain," because all sensory signals except olfaction must pass through it to get to the cortex. The cortex also projects profusely back to the thalamus; for many thalamic nuclei, these downward projections outnumber the upward ones by an order of magnitude. Most nuclei of the thalamus are *specific* nuclei, each of which connects to a relatively small area of cortex. There are also several *nonspecific* nuclei (including the reticular nucleus and the intralaminar nuclei), which extend diffuse, modulatory projections across most of the cortex—with a single axon forming synapses in many distinct areas—and receive projections from a similarly broad swath.

The broad connectivity of the thalamus and its central role in sensation have made it a promising target for neural theories of consciousness. One of the earliest such was Francis Crick's "thalamic searchlight hypothesis" (Crick, 1984), on which the thalamus controls which areas of cortex become the focus of consciousness. Since then, so-called *thalamocortical loop models* have been widely pursued, and this circuit now plays a role in many neural theories of consciousness; here we will focus on the version developed by Rodolfo Llinás. During the 1990s, Llinás and his coworkers conducted a series of detailed studies of thalamocortical interactions, and out of this work he has developed a theory that integrates data from waking and sleeping consciousness, addresses the binding problem, and provides a criterion for discriminating representations that can fill the hole vacated by the 40-Hz hypothesis (as discussed earlier).

First, it is important to understand how thalamocortical models in general account for binding. The common thread in these accounts is that thalamocortical interactions are necessary for the fast and precise generation of synchronous oscillations across distinct cortical regions.[4] In this respect the thalamus acts something like the conductor of a cortical symphony: It does not determine in detail what the players do, but it coordinates their activity and imposes coherence (Wang, Spencer, Fellous, & Sejnowski, 2010). Without this contribution from the thalamus, the brain might be able to produce local patches of synchrony, but it would not be able to bind the many different properties of a percept into a single coherent object.[5]

The problem of representing multiple separate bound objects at the same time could, in theory, be solved at least in part by ensuring that each bound representation oscillates at a different frequency. But this still leaves open the question of what distinguishes *conscious* bound representations. What determines which of several synchronously oscillating clusters dominates a person's subjective awareness?

Llinás has identified a mechanism that may subserve this function. Using MEG in humans, he has observed waves of phase-locked activity that "scan" across the cortex from the front of the head to the back. Each wave takes approximately 12.5 msec to traverse the brain, and is followed by a similar gap before the next wave, for a

---

[4] This represents a minimal necessary function for the thalamus that most models of this genre would agree on. There are many more specific questions on which accounts vary: For example, it is not clear how crucial the thalamus is for maintaining synchrony among neurons *within* a single cortical area, and though some neurons will oscillate even *in vitro* there is much debate about the extent to which oscillations observed in cortex derive from such "endogenous" oscillatory properties as opposed to system-level interactions.

[5] This metaphor also helps to illustrate why it is inaccurate to describe any individual part of the brain as the "seat of consciousness." A conductor and orchestra work together to produce coherent music; the conductor imposes structure on the orchestra, but in the end it is the individual musicians who produce the actual music. Likewise, the thalamus in some sense generates and directs consciousness, but only in conjunction with sensory areas that produce and embody the experienced *content* of that consciousness.

total interval of 25 ms per wave, or 40 Hz. Their presence is correlated with coherent conscious experience: They occur continuously during waking and REM sleep but vanish during nREM. These waves are apparently driven by the nonspecific nuclei of the thalamus, which send out projections that traverse the cortex from front to back (Llinás & Pare, 1996).

Llinás's hypothesis is that consciousness is marked by a second type of synchrony: synchrony between an individual cluster and this nonspecific scanning wave. Thus, of all the clusters that are active at a given time, the ones that are the focus of consciousness will be those that are oscillating in phase with the scanning wave.

A crucial line of evidence for this comes from Llinás's studies of auditory perception in humans during waking, REM, and slow-wave sleep (Llinás & Ribary, 1994). In awake humans, a salient auditory stimulus (a loud click) will interrupt the scanning wave and start a new one, while in REM the stimulus will produce a cortical response but will not reset the scanning wave. This would seem to correspond to the ability of such stimuli to draw conscious attention during waking but not during REM sleep (or during nREM, where the scanning wave is absent or at least dramatically reduced).

Another set of studies (Joliot et al., 1994) showed a different sort of correlation between this "gamma reset" and conscious perception. Subjects were played a pair of clicks separated by an interval between 3 and 30 ms. Subjects were able to distinguish the two clicks when they were separated by approximately 13 ms or more, but with shorter intervals they perceived only one click (of normal, not double, duration). MEG revealed that intervals under 12 ms produced only a single reset, whereas longer intervals produced two. The authors conclude from these results that consciousness is discrete rather than continuous, with 12 ms being the "quantum of consciousness," the basic temporal unit of conscious experience. Even for the more conservatively inclined, however, these two lines of evidence do strongly suggest that there is *some* close relationship between the scanning wave and conscious experience.

Gerald Edelman and Giulio Tononi (Edelman & Tononi, 2000) also emphasize the thalamocortical system, though their concern is less with synchrony itself than with the functional integration that it signifies. On their model, conscious neural representation is distinguished primarily by two characteristics: *integration*, the tendency of neurons within a particular representational cluster to interact more strongly with each other than with neurons outside the cluster; and *complexity*, the ability of the brain to select one specific state from a vast repertoire of possible states (and to do so several times a second). They use the term "dynamic core" to refer to a functional grouping of neurons that plays this role. The word *dynamic* is crucial here: For Edelman and Tononi (as for Llinás), the "core" of consciousness is not a persistent anatomical structure but an ephemeral pattern of activity that will be present in different areas of cortex (and different neurons within those areas) at different times. More recently, Tononi (2010) has also developed a more explicitly mathematical account of information integration, with the aim of developing a formal criterion of consciousness in neural systems.

Visceral Awareness and Self-Representation

Another major development in the study of consciousness has been the increasing degree of attention paid to the role of self-representation. Within philosophy, consciousness has often been analyzed in terms of a relation between transient mental objects or events—thoughts, ideas, sensations—and a persistent, unitary self. This approach has now been carried over into the empirical realm by neuroscientists, who are trying to determine how the brain represents the self and how this self-representation contributes to conscious experience.[6]

In *Descartes' Error*, Damasio (1994) defended the idea that conscious thought is substantially dependent on visceral self-perception. In his view, conscious decision making involves not only abstract reasoning but also the constant monitoring of a "body loop" in which brain and body respond to each other: Physiological mechanisms such as the endocrine system and sympathetic and parasympathetic nervous systems respond to external and internal stimuli that are represented by the brain, and the brainstem monitors the body and registers the state changes wrought by these systems. This gives literal meaning to the notion of a "gut instinct"; in numerous studies (Bechara, Damasio, Damasio, & Anderson, 1994; Damasio, 1996), Damasio and his co-workers have shown that physiological responses may be necessary for accurate decision making and may even register the "correct" answer to a problem before the subject is consciously aware of it.

---

[6] This is another point on which the convergence among major neural theories of consciousness is quite striking. Though we will focus on the work of Damasio, self-perception and its relation to decision making are accorded a central role in Edelman and Tononi 2000 and Llinás 2001.

Subsequently, Damasio (1999) has extended this model to provide an account of perceptual consciousness. Visceral self-representation now constitutes the "proto-self," a moment-to-moment sense of the presence and state of one's body. Mere perception becomes conscious experience when it is somehow integrated with or related to this proto-self, by way of second-order representations that register the body's response to a percept. "Core consciousness" is the realm of primary conscious experience, constituted by a series of these "how do I feel about what I'm seeing?" representations, and "extended consciousness" is the extension of these experiences into the past and future via the powers of memory and conceptual abstraction.

Damasio offers specific hypotheses about the neural localization of these functions. He believes that the self-representations that constitute the proto-self are generated by a number of upper brainstem structures (including much of what is traditionally referred to as the reticular system), the hypothalamus, and cortical somatosensory areas (primarily in right parietal cortex). Core consciousness depends primarily on the cingulate cortices and on the intralaminar (nonspecific) nuclei of the thalamus, and extended consciousness relies on the temporal and prefrontal cortices.

To interpret these claims, however, it is important to understand the particular notion of localization that Damasio is working with. He is a clinical neurologist, and his primary source of evidence is observation of humans with focal brain damage. Within the tradition of clinical neurology, the claim that "function F is localized to system S" rarely means more than "damage to system S will (more or less selectively) impair function F." In any case, this is the strongest claim that lesion data alone can usually justify. This restricted kind of localization is important, but it is also fundamentally incomplete as an *explanation* of the function in question, because it does not describe the *mechanism* by which the function is performed.[7] A theory of consciousness constructed along these lines can still have important consequences: For example, it guides us in interpreting the nature and subjective character of a range of neural pathologies, from Alzheimer's disease to locked-in syndrome, and it may help to establish the parameters for more focused study of individual functions. But outside of the diagnostic realm, its utility will be limited unless and until it can be supplemented with the sort of mechanistic underpinning that supports more fine-grained prediction, testing, and explanation.

CONCLUSION: THE FUTURE OF CONSCIOUSNESS

The mind–body problem and many of the problems encountered in the study of consciousness may result from the separate mental models (or conceptual schemes) we use to think about mental events and physical events. Mental models influence our thinking profoundly, providing the structure within which we frame problems and evaluate solutions. However, it remains possible to distinguish properties of the model from properties of reality. As it stands, our models of the mental and the physical are distinct, but this may be more a symptom of our flawed understanding than a fact about the world itself.

One way to understand the progress described in this chapter is as a breaking down of this dualist divide. Psychologists studying consciousness have found ways to relate it to the physical behavior of the organism, forging epistemological links between mind and world. In addition, as we have learned more about the detailed structure of mental functions such as attention, perception, memory, and decision making, it has become less and less tempting to see them as parts of a transcendent consciousness. Meanwhile, neuroscience has begun to elucidate the ontological connections between mind and body, making it possible to see where our models of the mental and physical may overlap and eventually merge. Though such a full integration may be far off, a path to the goal is now in view; rather than positing mere *correlations* between mind and brain, the best modern theories draw on rich, detailed parallels in both structure and function, permitting precise, testable predictions and causal interventions that span the mind–body divide from both directions (Farber, 2005).

These developments cause us to reflect with some amazement on the history of the scientific study of consciousness. Until the ascendency of behaviorism in the early part of the 20th century it was widely considered to be the central object of the field of psychology. Then, through the behaviorist era and until late in the cognitive

---

[7] By way of illustration, consider the following statements: (1) "The lungs are the organs that oxygenate the blood." (2) "The lungs contain a honeycomb of air vessels, and hence have a very high internal surface area. Blood is pumped directly through the lungs and is brought to the surface of these vessels, allowing for the exchange of gases with the inhaled air." Both are in some sense localizations of the function of oxygenation, but the first explains nothing about the *means* by which the function is performed. For this very reason, it is also easier to formulate and confirm—for example, by measuring the oxygen content of blood flowing into and out of the lungs.

revolution, which began in the 1960s, it was banished entirely from study. Now, it may provide a new center to integrate the diverse areas of cognition and help relate them to dramatic new findings from neuroscience.

What can be said about the future of consciousness? There is an instructive parallel here with the history of life (Farber, 2005). At the turn of the last century, there was still room for doubt as to whether there would ever be a unified account of living and nonliving processes. As with consciousness, there was (and to a certain extent, still is) a division between the mental models we use to describe the behavior of animate and inanimate objects. Vitalists argued that this division was reflected in reality, while materialists argued that life was ultimately grounded in the same physical forces and entities as everything else. As it turned out, the vitalists were wrong, and the elaboration of the physical basis of life revolutionized biology and led directly to many of the greatest scientific advances of the 20th century.

We cannot say with certainty whether psychological materialism will enjoy a similar victory; the current rate of progress certainly gives grounds for optimism, but there are also many deep conceptual problems that remain to be overcome. The real value of the parallel with the history of life is not in prediction, but in understanding the nature of the problem and how best to approach it. Vitalists posed deep philosophical problems having to do with seemingly unique properties of living organisms, such as self-reproduction and goal-directedness. Progress came neither from ignoring these problems nor from accepting them on their own terms, but from *reinterpreting* them as challenging scientific puzzles, puzzles that could only be solved with a combination of empirical and theoretical advances.

It is also important to notice that the victory of biological materialism did not lead to the concept of life being discarded or biology becoming a branch of physics. The word *life* is still available for everyday use and remains just as respectable as it was a hundred years ago, even though its meaning now depends in part on the scientific explanation that has developed during that time. Because the word *consciousness* has always been more obscure and more diverse in its meanings, it may be in for somewhat more radical change; scientists working on consciousness may rely on more technical terms (such as *attention, awareness,* and *binding*), just as biologists conduct much of their daily work without general reference to *life*; but there is no reason to presume that scientific progress will involve rejecting the very idea of consciousness or replacing mental terms with behavioral or neural ones. Explaining need not mean explaining away.

REFERENCES

Adams, J. K. (1957). Laboratory studies of behavior without awareness. *Psychological Bulletin*, *54,* 383–405.

Aristotle. (1991). *De anima* (R. D. Hicks, Trans.) Buffalo, NY: Prometheus Books.

Baars, B. J. (1988). *A cognitive theory of consciousness*. New York, NY: Cambridge University Press.

Baars, B. J. (1997). *In the theater of consciousness: The workspace of the mind*. New York, NY: Oxford University Press.

Baars, B. J., Banks, W. P., & Newman, J. B. (Eds.) (2003). *Essential sources in the scientific study of consciousness*. Cambridge, MA: MIT Press.

Baldwin, M. W. (2001) Relational schema activation: Does Bob Zajonc ever scowl at you from the back of your mind? In J. A. Bargh, & D. K. Apsley (Eds.), *Unraveling the complexities of social life: A festschrift in honor of Robert B. Zajonc* (pp. 55–67). Washington, DC: American Psychological Association.

Banks, W. P. (1993). Problems in the scientific pursuit of consciousness. *Consciousness & Cognition: An International Journal*, *2,* 255–263.

Banks, W. P., Roberts, D., & Ciranni, M. (1995). Negative priming in auditory attention. *Journal of Experimental Psychology: Human Perception & Performance*, *21,* 1354–1361.

Barber, T. X. (2000). A deeper understanding of hypnosis: Its secrets, its nature, its essence. *American Journal of Clinical Hypnosis*, *42,* 208–272.

Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*, *50,* 7–15.

Bisiach, E., Cornacchia, L., Sterzi, R., & Vallar, G. (1984). Disorders of perceived auditory lateralization after lesions of the right hemisphere. *Brain*, *107* (Pt 1), 37–52.

Bisiach, E., & Geminiani, G. (1991). Anosognosia related to hemiplegia and hemianopia. In G. P. Prigatano & D. L. Schacter (Eds.), *Awareness of deficit after brain injury: Clinical and theoretical issues* (pp. 17–39). New York, NY: Oxford University Press.

Bisiach, E., & Luzzatti, C. (1978). Unilateral neglect of representational space. *Cortex*, *14,* 129–133.

Bisiach, E., & Rusconi, M. L. (1990). Break-down of perceptual awareness in unilateral neglect. *Cortex*, *26,* 643–649.

Boring, E. G. (1942). *Sensation and perception in the history of experimental psychology*. New York, NY: Appleton.

Bornstein, R. F., & Pittman, T. S. (Eds.). (1992). *Perception without awareness: Cognitive, clinical, and social perspectives*. New York, NY: Guilford Press.

Broadbent, D. E. (1958). *Perception and communication*. New York: Pergamon Press.

Bruner, J. (1992). Another look at new look 1. *American Psychologist*, *47,* 780–783.

Bunce, S. C., Bernat, E., Wong, P. S., & Shevrin, H. (1999). Further evidence for unconscious learning: Preliminary support for the conditioning of facial EMG to subliminal stimuli. *Journal of Psychiatric Research*, *33,* 341–347.

Chalmers, D. J. (1996). *The conscious mind: In search of a fundamental theory*. New York, NY: Oxford University Press.

Chan, J. L., & Liu, A. B. (1999) Anatomical correlates of alien hand syndromes. *Neuropsychiatry Neuropsychology & Behavioral Neurology*, *12,* 149–155.

Cheesman, J., & Merikle, P. M. (1984). Priming with and without awareness. *Perception & Psychophysics*, *36,* 387–395.

Cheesman, J., & Merikle, P. M. (1986). Distinguishing conscious from unconscious perceptual processes. *Canadian Journal of Psychology*, *40,* 343–367.

Chen, Z., & Treisman, A. (2009). Implicit perception and level of processing in object-substitution masking. *Psychological Science*, *20,* 560–567.

Cherry, C. (1957). *On human communication: a review, a survey, and a criticism*. Cambridge, MA: The Technology Press of MIT.

Chomsky, N. (1996). A review of B. F. Skinner's *Verbal Behavior*. In H. Geirsson & M. Losonsky (Eds.). *Readings in language and mind* (pp. 413–441). Cambridge, MA: Blackwell.

Churchland, P. S. (1986). *Neurophilosophy: Toward a unified science of the mind-brain*. Cambridge, MA: MIT Press.

Churchland, P. S. (1996). The hornswoggle problem. *Journal of Consciousness Studies*, *3,* 402–408.

Cohen, N. J., Eichenbaum, H., Deacedo, B. S., & Corkin, S. (1985). Different memory systems underlying acquisition of procedural and declarative knowledge. *Annals of the New York Academy of Sciences*, *444,* 54–71.

Cooke, P. (1991, June 23). They cried until they could not see. *New York Times Magazine*, 46–47.

Corteen, R. S. (1986). Electrodermal responses to words in an irrelevant message: A partial reappraisal. *Behavioral and Brain Sciences*, *9,* 27–28.

Corteen, R. S., & Dunn, D. (1974). Shock-associated words in a nonattended message: A test for momentary awareness. *Journal of Experimental Psychology*, *102,* 1143–1144.

Corteen, R. S., & Wood, B. (1972). Autonomic responses to shock-associated words in an unattended channel. *Journal of Experimental Psychology*, *94,* 308–313.

Cowan, N., & Wood, N. L. (1997). Constraints on awareness, attention, processing, and memory: Some recent investigations with ignored speech. *Consciousness and Cognition*, *6,* 182–203.

Crick, F. (1984). Function of the thalamic reticular complex: The searchlight hypothesis. *Proceedings of the National Academy of Science*, *81,* 4586–4590.

Crick, F., & Koch, C. (1995). Are we aware of neural activity in primary visual cortex? *Nature*, *375* (6527), 121–123.

Crick, F., & Koch, C. (1998). Constraints on cortical and thalamic projections: The no-strong-loops hypothesis. *Nature*, *391* (6664), 245–250.

Crick, F., & Koch, C. (2003). A framework for consciousness. *Nature Neuroscience*, *6,* 119–126.

Dalrymple-Alford, E. C., & Budayr, B. (1966). Examination of some aspects of the Stroop color-word test. *Perceptual & Motor Skills*, *23,* 1211–1214.

Damasio, A. R. (1994). *Descartes' error: Emotion, reason, and the human brain*. New York, NY: Putnam.

Damasio, A. R. (1996). The somatic marker hypothesis and the possible functions of the prefrontal cortex. *Philosophical Transactions of the Royal Society of London B Biological Science*, *351,* 1413–1420.

Damasio, A. R. (1999). *The feeling of what happens: Body and emotion in the making of consciousness* (1st ed.). New York, NY: Harcourt Brace.

Dardis, A. (2008). *Mental causation: The mind-body problem*. New York, NY: Columbia University Press.

Dawson, M. E., & Schell, A. M. (1982). Electrodermal responses to attended and nonattended significant stimuli during dichotic listening. *Journal of Experimental Psychology: Human Perception and Performance*, *8,* 315–324.

Day, R. H., & Kimm, A. C. (2010). Analysis and explanation of the Thiéry-Wundt Illusion. *Perception*, *39,* 942–952.

Debner, J. A., & Jacoby, L. L. (1994). Unconscious perception: Attention, awareness, and control. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *20,* 304–317.

de Gelder, B., Vroomen, J., Pourtois, G., & Weiskrantz, L. (1999). Non-conscious recognition of affect in the absence of striate cortex. *Neuroreport*, *10,* 3759–3763.

Dennett, D. C. (1991). *Consciousness explained*. Boston, MA: Little Brown.

Descartes, R. (1951). *Meditations* (Laurence J. Lafleur, Trans.) New York, NY: Liberal Arts Press. (Original work published 1641)

Deutsch, J. A., & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review*, *70,* 80–90.

Dixon, N. F. (1971). *Subliminal perception: The nature of a controversy*. London: McGraw-Hill.

Dixon, N. F. (1981) *Preconscious processing*. London: Wiley.

Draine, S. C., & Greenwald, A. G. (1998). Replicable unconscious semantic priming. *Journal of Experimental Psychology: General*, *127,* 286–303.

Dzhafarov, E. N., & Colonius, H. (2011). The Fechnerian idea. *The American Journal of Psychology*, *124,* 127–140.

Ebbinghaus, H. (1908). *Psychology: An elementary text-book* (M. Meyer, Trans.) Boston, MA: Heath.

Edelman, G. M., & Tononi, G. (2000). *A universe of consciousness: How matter becomes imagination*. New York, NY: Basic Books.

Engel, A. K., Fries, P., Konig, P., Brecht, M., & Singer, W. (1999). Temporal binding, binocular rivalry, and consciousness. *Consciousness and Cognition*, *8,* 128–151.

Engel, A. K., & Singer, W. (2001). Temporal binding and the neural correlates of sensory awareness. *Trends in Cognitive Science*, *5,* 16–25.

Erdelyi, M. H. (1972). Role of fantasy in the Poetzl (emergence) phenomenon. *Journal of Personality and Social Psychology*, *24,* 186–190.

Erdelyi, M. H. (1974). A new look at the new look: Perceptual defense and vigilance. *Psychological Review*, *81,* 1–25.

Eriksen, C. W. (1958). Unconscious processes. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1958* (pp. 169–278). Lincoln: University of Nebraska Press.

Eriksen, C. W. (1960). Discrimination and learning without awareness: A methodological survey and evaluation. *Psychological Review*, *67,* 279–300.

Eriksen, C. W. (Ed.). (1962). Behavior and awareness: A symposium of research and interpretation. *Journal of Personality*, *30,* 158.

Farber, I. (2005). How a neural correlate can function as an explanation of consciousness: Evidence from the history of science regarding the likely explanatory value of the NCC approach. *Journal of Consciousness Studies*, *12* (4-5), 77–95.

Farber, I., & Churchland, P. S. (1995). Consciousness and the neurosciences: Philosophical and theoretical issues. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 668–688). Cambridge, MA: MIT Press.

Festinger, L., & Easton, A. M. (1974). Inferences about the efferent system based on a perceptual illusion produced by eye movements. *Psychological Review*, *81,* 44–58.

Forster, P. M., & Govier, E. (1978) Discrimination without awareness? *Quarterly Journal of Experimental Psychology*, *30,* 289–295.

Franklin, S., & Graesser, A. (1999). A software agent model of consciousness. *Consciousness & Cognition*, *8,* 285–301.

Freud, S. (1965). *The psychopathology of everyday life* (A. Tyson, Trans.) New York, NY: Norton.

Frith, C. D. (1992). Consciousness, information processing and the brain. *Journal of Psychopharmacology*, *6,* 436–440.

Frith, C. D. (1996). The role of the prefrontal cortex in self-consciousness: The case of auditory hallucinations. *Philosophical Transactions of the Royal Society: Biological Sciences*, *351,* 1505–1512.

Fuhrer, M. J., & Eriksen, C. W. (1960). The unconscious perception of the meaning of verbal stimuli. *Journal of Abnormal and Social Psychology*, *61,* 432–439.

Galin, D. (1992). Theoretical reflections on awareness, monitoring, and self in relation to anosognosia. *Consciousness & Cognition*, *1,* 152–162.

Gardner, H. (1985). *The mind's new science: A history of the cognitive revolution*. New York, NY: Basic Books.

Gazzaniga, M. S. (2005). Forty-five years of split-brain research and still going strong. *Nature Reviews Neuroscience*, *6,* 653–659.

Greenfield, S. (1995). *Journey to the centers of the mind: Toward a science of consciousness*. New York, NY: W. H. Freeman.

Greenwald, A. G. (1992). New Look 3: Unconscious cognition reclaimed. *American Psychologist*, *47,* 766–779.

Greenwald, A. G., Klinger, M. R., & Schuh, E. S. (1995). Activation by marginally perceptible ("subliminal") stimuli: Dissociation of unconscious from conscious cognition. *Journal of Experimental Psychology: General*, *124,* 22–42.

Heilman, K. M., Barrett, A. M., & Adair, J. C. (1998). Possible mechanisms of anosognosia: A defect in self-awareness. *Philosophical Transactions of the Royal Society: Biological Sciences*, *353,* 1903–1909.

Henkel, L. A., Franklin, N., & Johnson, M. K. (2000). Cross-modal source monitoring confusions between perceived and imagined events. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26,* 321–335.

Hilgard, E. R. (1977). *Divided consciousness: Multiple controls in human thought and action*. New York, NY: Wiley.

Hilgard, E. R. (1992). Divided consciousness and dissociation. *Consciousness & Cognition*, *1,* 6–31.

Hill, C. S. (2009). *Consciousness*. New York, NY: Cambridge University Press.

Holender, D. (1986) Semantic activation without conscious identification in dichotic listening, parafoveal vision, and visual masking: A survey and appraisal. *Behavioral & Brain Sciences*, *9,* 1–66.

Inzelberg, R., Nisipeanu, P., Blumen, S. C., & Carasso, R. L. (2000). Alien hand sign in Creutzfeldt-Jakob disease. *Journal of Neurology, Neurosurgery & Psychiatry*, *68,* January, 103–104.

Ionescu, M. D., & Erdelyi, M. H. (1992). The direct recovery of subliminal stimuli. In R. F. Bornstein, T. S. Pittman, & S. Thane (Eds.), *Perception without awareness: Cognitive, clinical, and social perspectives* (pp. 143–169). New York, NY: Guilford Press.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory & Language*, *30,* 513–541.

Jacoby, L. L., Lindsay, D. S., & Toth, J. P. (1992). Unconscious influences revealed: Attention, awareness, and control. *American Psychologist*, *47,* 802–809.

Jacoby, L. L., Toth, J. P., & Yonelinas, A. P. (1993). Separating conscious and unconscious influences of memory: Measuring recollection. *Journal of Experimental Psychology: General*, *122,* 139–154.

James, W. (1983). *The Principles of Psychology*. Reprint of 1890 edition. Cambridge, MA: Harvard University Press.

Joliot, M., Ribary, U., & Llinás, R. (1994). Human oscillatory brain activity near 40 Hz coexists with cognitive temporal binding. *Proceedings of the National Academy of Science*, *91,* 11748–11751.

Kerkhoff, G. (2001). Spatial hemineglect in humans. *Progress in Neurobiology*, *63,* 1–27.

Kihlstrom, J. F. (1987). The cognitive unconscious. *Science*, *237,* 1445–1452.

Kihlstrom, J. F. (1992). Dissociation and dissociations: A comment on consciousness and cognition. *Consciousness & Cognition*, *1,* 47–53.

Kihlstrom, J. F. (1994). Hypnosis, delayed recall, and the principles of memory. *International Journal of Clinical & Experimental Hypnosis*, *42,* 337–345.

Kihlstrom, J. F. (1997). Consciousness and me-ness. In J. D. Cohen & J. W. Schooler (Eds.), *Scientific approaches to consciousness* (pp. 451–468). Mahwah, NJ: Erlbaum.

Kihlstrom, J. F., Barnhardt, T. M., & Tataryn, D. J. (1992). Implicit perception. In R. Bornstein & T. S. Pittman (Eds.), *Perception without awareness: Cognitive, clinical, and social perspectives* (pp. 17–54). New York, NY: Guilford Press.

Koch, C. (2004). *The quest for consciousness: A neurobiological approach*. Englewood, CO: Roberts.

Koch, C., & Crick, F. (1994). Some further ideas regarding the neuronal basis of awareness. In C. Koch & J. L. Davis (Eds.), *Large-scale neuronal theories of the brain*. Cambridge, MA: MIT Press.

Koch, C., & Davis, J. L. (1994). *Large-scale neuronal theories of the brain*. Cambridge, MA: MIT Press.

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago, IL: University of Chicago Press.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago, IL: University of Chicago Press.

Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral & Brain Sciences*, *8,* 529–566.

Libet, B. (1993). The neural time factor in conscious and unconscious events. In *Experimental and theoretical studies of consciousness*. Ciba Foundation Symposium, 174 (pp. 123–146). Chichester, England: Wiley.

Libet, B., Alberts, W. W., & Wright, E. W., Jr. (1967). Responses of human somatosensory cortex to stimuli below threshold for conscious sensation. *Science*, *158,* 1597–1600.

Libet, B., Alberts, W. W., Wright, E. W., Delattre, L. D., Levin, G., & Feinstein, B. (1964). Production of threshold levels of conscious sensation by electrical stimulation of human somatosensory cortex. *Journal of Neurophysiology*, *27,* 546–578.

Llinás, R. R. (2001). *I of the vortex: From neurons to self* . Cambridge MA: MIT Press.

Llinás, R., & Pare, D. (1996). The brain as a closed system modulated by the senses. In R. Llinás & P. S. Churchland (Eds.), *The mind-brain continuum: Sensory processes*. Cambridge, MA: MIT Press.

Llinás, R., & Ribary, U. (1994). Perception as an oneiric-like state modulated by the senses. In C. Koch & J. L. Davis (Eds.), *Large-scale neuronal theories of the brain*. Cambridge MA: MIT Press.

Llinás, R., Ribary, U., Contreras, D., & Pedroarena, C. (1998). The neuronal basis for consciousness. *Philosophical Transactions of the Royal Society: Biological Sciences*, *353,* 1841–1849.

Loftus, E. F. (1993) Psychologists in the eyewitness world. *American Psychologist*, *48,* 550–552.

Logothetis, N. K. (1998). Single units and conscious vision. *Philosophical Transactions of the Royal Society: Biological Sciences*, *353,* 1801–1818.

Logothetis, N. K., & Schall, J. D. (1989). Neuronal correlates of subjective visual perception. *Science*, *245,* 761–763.

Logothetis, N. K., & Sheinberg, D. L. (1996). Recognition and representation of visual objects in primates: Psychophysics and physiology. In R. Llinás, R. Rodolfo, & P. S. Churchland (Eds.), *The mind-brain continuum: Sensory processes* (pp. 147–172). Cambridge, MA: MIT Press.

MacDonald, P. A. & Joordens, S. (2000). Investigation a memory-based account of negative priming: Support for selection-feature mismatch. *Journal of Experimental Psychology: Human Perception & Performance*, *26,* 1478–1496.

Mack, A. & Rock, I. (1998). *Inattentional blindness*. Cambridge, MA: MIT Press.

Mackay, D. G. (1973). Aspects of the theory of comprehension, memory and attention. *Quarterly Journal of Experimental Psychology*, *25,* 22–40.

Marcel, A. J. (1983a). Conscious and unconscious perception: An approach to the relations between phenomenal experience and perceptual processes. *Cognitive Psychology*, *15,* 238–300.

Marcel, A. J. (1983b). Conscious and unconscious perception: Experiments on visual masking and word recognition. *Cognitive Psychology*, *15,* 197–237.

McCloskey, M. & Kohl, D. (1983). Naive physics: The curvilinear impetus principle and its role in interactions with moving objects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9,* 146–156.

McGinnies, E. (1949). Emotionality and perceptual defense. *Psychological Review*, *56,* 244–251.

Merikle, P. M., Joordens, S., & Stolz, J. A. (1995). Measuring the relative magnitude of unconscious influences. *Consciousness & Cognition*, *4,* 422–439.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, *90,* 227–234.

Milliken, B., Joordens, S., Merikle, P. M., & Seiffert, A. E. (1998). Selective attention: A reevaluation of the implications of negative priming. *Psychological Review*, *105,* 203–229.

Moran, J. & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science*, *229,* 782–784.

Moray, N. (1959). Attention in dichotic listening: Affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology*, *11,* 56–60.

Morris, J. S., DeGelder, B., Weiskrantz, L., & Dolan, R. J. (2001). Differential extrageniculostriate and amygdala responses to presentation of emotional faces in a cortically blind field. *Brain*, *124,* 1241–1252.

Moscovitch, M., & Melo, B. (1997). Strategic retrieval and the frontal lobes: Evidence from confabulation and amnesia. *Neuropsychologia*, *35,* 1017–1034.

Neisser, U. (1967). *Cognitive psychology*. New York, NY: Appleton-Century-Crofts.

Newstead, S. E., & Dennis, I. (1979). Lexical and grammatical processing of unshadowed messages: A re-examination of the Mackay effect. *Quarterly Journal of Experimental Psychology*, *31,* 477–488.

Nissen, M. J., Ross, J. L., Willingham, D. B., Mackenzie, T. B., & Schacter, D. L. (1994). Evaluating amnesia in multiple personality disorder. In R. M. Klein & B. K. Doane (Eds.), *Psychological concepts and dissociative disorders* (pp. 259–282). Hillsdale, NJ: Erlbaum.

Peirce, C. S., & Jastrow, J. (1884). On small differences in sensation. *Memoirs of the National Academy of Sciences*, *3,* 73–83.

Proctor, R. W., & Vu, K. L. (2006). The cognitive revolution at age 50: Has the promise of the human information-processing approach been fulfilled? *International Journal of Human-Computer Interaction*, *21,* 253–284.

Proust, J. (2000). Awareness of agency: Three levels of analysis. In T. Metzinger (Ed.), *Neural correlates of consciousness: Empirical and conceptual questions* (pp. 307–324). Cambridge, MA: MIT Press.

Ramachandran, V. S. (1995). Anosognosia in parietal lobe syndrome. *Consciousness & Cognition*, *4,* 22–51.

Ramachandran, V. S. (1996). The evolutionary biology of self-deception, laughter, dreaming and depression: Some clues from anosognosia. *Medical Hypotheses*, *47,* 347–362.

Ramachandran, V. S., & Blakeslee, S. (1998). *Phantoms in the brain: Probing the mysteries of the human mind*. New York, NY: William Morrow.

Rock, I. (1983). *The logic of perception*. Cambridge, MA: MIT Press.

Roediger, H. L., III, & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21,* 803–814.

Ryle, G. (1949). *The concept of mind*. New York, NY: Hutchinson's University Library.

Schacter, D. L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13,* 501–518.

Schacter, D. L. (1990). Toward a cognitive neuropsychology of awareness: Implicit knowledge and anosognosia. *Journal of Clinical and Experimental Neuropsychology*, *12,* 155–178.

Schacter, D. L. (1995). *Memory distortion: How minds, brains, and societies reconstruct the past*. Cambridge, MA: Harvard University Press.

Searle, J. R. (1992). *The rediscovery of the mind*. Cambridge, MA: MIT Press.

Searle, J. R. (1993). The problem of consciousness. *Consciousness and Cognition*, *2,* 310–319.

Shevrin, H. (2000). The experimental investigation of unconscious conflict, unconscious affect, and unconscious signal anxiety. In M. Velmans (Ed.), *Investigating phenomenal consciousness: New methodologies and maps* (pp. 33–65). Amsterdam: John Benjamins.

Shevrin, H. & Luborsky, L. (1961). The rebus technique: A method for studying primary-process transformations of briefly exposed pictures. *Journal of Nervous & Mental Disease*, *133,* 479–488.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review*, *84,* 127–190.

Sidis, B. (1898). *The psychology of suggestion*. New York, NY: Appleton.

Silverman, L. H. (1983). The subliminal psychodynamic method: Overview and comprehensive listing of studies. In J. Masling (Ed.), *Empirical studies of psychoanalytic theory* (Vol. 1). Hillsdale, NJ: Erlbaum.

Simons, D. J., Hannula, D. E., Warren, D. E., & Day, S. W. (2007). Behavioral, neuroimaging, and neuropsychological approaches to implicit perception. In P. Zelazo, M. Moscovitch, & E. Thompson (Eds.), *The Cambridge handbook of consciousness* (pp. 207–250). New York, NY: Cambridge University Press.

Singer, W. (1996). Neuronal synchronization: A solution to the binding problem? In R. Llinás & P. S. Churchland (Eds.), *The mind-brain continuum: Sensory processes*. Cambridge, MA: MIT Press.

Singer, W. (2001). Consciousness and the binding problem. *Annals of the New York Academy of Science*, *929,* 123–146.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18,* 643–662.

Tipper, S. P. (1985). The negative priming effect: Inhibitory priming by ignored objects. *Quarterly Journal of Experimental Psychology*, *37A,* 571–590.

Tononi, G. (2010) Information integration: Its relevance to brain function and consciousness. *Archives of Italian Biology*, *148,* 299–322.

Tononi, G. & Koch, C. (2008). The neural correlates of consciousness: An update. *Annals of the New York Academy of Sciences*, *1124,* 239–261.

Treisman, A. M. (1964). Selective attention in man. *British Medical Bulletin*, *20,* 12–16.

Treisman, A. M., & Geffen, G. (1967). Selective attention: Perception or response? *Quarterly Journal of Experimental Psychology*, *19,* 1–17.

Treisman, A. M., & Kanwisher, N. G. (1998). Perceiving visually presented objects: Recognition, awareness, and modularity. *Current Opinion in Neurobiology*, *8,* 218–226.

Von der Malsburg, C. (1981). *The correlation theory of brain function* (Internal Report 81-2.): Department of Neurobiology, Max Planck Institute for Biophysical Chemistry, Gottingen, Germany.

Von Wright, J. M., Anderson, K., & Stenman, U. (1975). Generalisation of conditioned GSR's in dichotic listening. In P. M. A. Rabbitt & S. Dornic (Eds.), *Attention and Performance V*. New York, NY: Academic Press.

Wang, H., Spencer D., Fellous, J., & Sejnowski, T. (2010). Synchrony of thalamocortical inputs maximizes cortical reliability. *Science*, *5974,* 106–109.

Wardlaw, K. A., & Kroll, N. E. (1976). Autonomic responses to shock-associated words in a nonattended message: A failure to replicate. *Journal of Experimental Psychology: Human Perception and Performance*, *2,* 357–360.

Watson, J. B. (1925). *Behaviorism*. New York, NY: People's Institute.

Wegner, D. M. (1994). Ironic processes of mental control. *Psychological Review*, *101,* 34–52.

Weinberger, J. (1992). Validating and demystifying subliminal psychodynamic activation. In R. F. Bornstein & T. S. Pittman (Eds.), *Perception without awareness: Cognitive, clinical, and social perspectives* (pp. 170–188). New York, NY: Guilford Press.

Weiskrantz, L. (1990). (1989). Outlooks for blindsight: Explicit methodologies for implicit processes. *Proceedings of the Royal Society of London B Biological Sciences*, *239,* 247–278.

Weiskrantz, L. (1998). Pupillary responses with and without awareness in blindsight. *Consciousness & Cognition*, *7,* 324–326.

Welshon, R. (2011). *Philosophy, neuroscience and consciousness*. Montreal, Canada: McGill-Queen's University Press.

Wittgenstein, L. (1953). *Philosophical investigations*; translated by G. E. M. Anscombe. New York, NY: Macmillan.

Wong, P. S., Bernat, E., Bunce, S., & Shevrin, H. (1997). Brain indices of nonconscious associative learning. *Consciousness & Cognition*, *6,* 519–544.

Wood, N. L., & Cowan, N. (1995a). The cocktail party phenomenon revisited: Attention and memory in the classic selective listening procedure of Cherry (1953). *Journal of Experimental Psychology: General*, *124,* 243–262.

Wood, N. L., & Cowan, N. (1995b). The cocktail party phenomenon revisited: How frequent are attention shifts to one's name in an irrelevant auditory channel? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21,* 255–260.