The Neurobiology of Consciousness

CHRISTOF KOCH

ABSTRACT: Over the past two decades, the mystery of consciousness and its material basis has begun to be investigated by the sciences. The neurobiological approach to consciousness aims to identify its correlates at the neuronal level. Electrophysiological, psychophysical, and functional imaging studies in humans and nonhuman animals have allowed brain scientists to narrow their focus to the neural substrates of consciousness and conscious perception in the circuits of the forebrain, in particular the thalamocortical system and its satellites. These findings, complemented by the development of a robust theoretical predictive framework, should eventually lead to a rational understanding of the phenomenon of consciousness.

Consciousness is one of the most enigmatic features of the universe. People not only act but feel: they see, hear, smell, recall, plan for the future. These activities are associated with subjective, ineffable, immaterial feelings that are tied in some manner to the material brain. The exact nature of this relationship—the classical mind-body problem—remains elusive and the subject of heated debate (see Block, chapter 77, this volume). These firsthand, subjective experiences pose a daunting challenge to the scientific method that has, in many other areas, proven so immensely fruitful. The brute fact of consciousness comes as a total surprise; it does not appear to follow from any phenomena in traditional physics or biology.

People willingly concede that when it comes to nuclear physics or molecular biology, specialist knowledge is essential; but many assume that there are few relevant facts about consciousness, and therefore everybody is entitled to his or her own theory. Nothing could be further from the truth. There is an immense amount of relevant psychological, clinical, and neuroscientific data and observations that need to be accounted for.

Consciousness is a state-dependent property of certain complex, biological, adaptive, and highly interconnected systems. The best example of consciousness is found in a healthy and attentive human brain. In deep sleep, consciousness ceases. Small lesions in the midbrain, brain stem, and thalamus can lead to a complete loss of consciousness—probably through inactivation of the cerebral cortex (Sukhotinsky et al., 2007)—while destruction of circumscribed parts of the cerebral cortex can eliminate very specific aspects of consciousness, such as the ability to be aware of motion or to recognize faces, without a concomitant loss of vision in general. Brain scientists are focusing on experimental approaches that shed light on the neural basis of consciousness rather than on eristic philosophical arguments with no clear resolution. This chapter reviews these experimental approaches.

What phenomena does consciousness encompass?

Consciousness has been dissected on conceptual grounds (access versus phenomenal consciousness; see Block, 2005, and chapter 77, this volume), ontological grounds (Hard versus Easy problem; Chalmers, 1996), and psychological grounds (explicit versus implicit processes; Tulving, 1993).

One common philosophical definition is “Consciousness is what it is like to be something,” such as the experience of what it feels like to smell a rose or to be in love. This what-it-feels-like-from-within definition expresses the principal irreducible characteristic of the phenomenal aspect of consciousness: to experience something.

What it feels like to have a particular experience is called the quale of that experience: the quale of red is what is common to such disparate conscious states as seeing a red sunset, the red flag of China, arterial blood, or a ruby gemstone. All four subjective states share “redness.” There are countless qualia (the plural of quale): the ways things look, sound, and smell, the way it feels to have a pain, the way it feels to have thoughts and desires, and so on. To have an experience means to have qualia, and the quale of an experience is what specifies it and makes it different from other experiences.

Science must explain the exact relationship between the immaterial, conscious mind and its physical basis in the electrochemical interactions in the body. This challenge can be decomposed into several subproblems.

Why is there any experience at all? Why does a brain state feel like anything? Many scholars have argued that the exact nature of this relationship will remain a central puzzle of
human existence, without an adequate reductionistic, scientific explanation. However, similar sentiments have been expressed in the past for the problem of seeking to understand life or to determine what material the stars are made of. Thus it is best to put this question aside for the moment and not be taken in by defeatist arguments.

**Why is the relationship among different experiences the way it is?** For instance, red, yellow, green, cyan, blue, and magenta are all colors that can be mapped onto the topology of a circle. Why? Furthermore, as a group, these color percepts share certain communalities that make them different from other percepts, such as seeing motion or smelling a rose. Why?

**Why are feelings private?** As expressed by poets and novelists, we cannot communicate an experience to somebody else except by way of example.

**How do feelings acquire meaning?** Subjective states are not abstract states but have an immense amount of associated explicit and implicit feelings. Think of the unmistakable smell of dogs coming in from the rain or the crunchy texture of potato chips. How do these arise?

**Why are only some behaviors associated with conscious states?** Much brain activity and many associated behaviors occur without any conscious sensation. Why? And where is the difference between the two at the neuronal level?

**The neurobiology of free will**

A further aspect of the mind-body problem is the question of free will and will power. Answering this question goes to the heart of the way people think of themselves. Of great relevance are the classical findings by Libet and colleagues (1983) of brain events that precede the conscious initiation of a voluntary action. This simple result has been replicated and extended but, because of its counterintuitive implication that conscious will has no causal role, continues to be vigorously debated (Haggard & Eimer, 1999; Soon, Brass, Heinze, & Haynes, 2008; Brass & Haggard, 2008; see Lau, chapter 83, this volume).

Psychological work in both normals and patients reveals dissociations between the conscious perception of a willed action and its actual execution: subjects believe that they perform actions that they did not do while, under different circumstances, subjects feel that they are not responsible for actions that are, demonstrably, their own (Wegner, 2002).

Whether volition is illusory or is free in some libertarian sense does not answer the question of how subjective states relate to brain states. The perception of free will, which psychologists call the feeling of agency or authorship (e.g., “I decided to lift my finger”), is a subjective state with an associated quale no different in kind from the quale of a toothache or seeing marine blue. It must also have some neuronal correlate. Direct electrical brain stimulation during neurosurgery, as well as fMRI experiments, implicates medial premotor and anterior cingulate cortices in generating the subjective feeling of triggering an action (see Lau, chapter 83, this volume).

**Consciousness in other species**

Data about conscious states come not only from people who can talk about their subjective experiences but also from nonlinguistic competent individuals—newborn babies (Lagerctantz & Changeux, 2009) or patients with complete paralysis of nearly all voluntary muscles (locked-in syndrome)—and, most importantly, from animals other than humans. There are at least three reasons to assume that many species, in particular those with complex behaviors such as mammals, share at least some aspects of consciousness with humans:

**Similar neuronal architectures** Except for size, there are no large-scale, dramatic differences between the cerebral cortex and thalamus of mice, monkeys, humans, and whales. It is difficult to distinguish a cubic millimeter of neocortex among different mammals, except by expert neuroanatomists.

**Similar behavior** Almost all human behaviors have precursors in the animal literature. Take the case of pain. The behaviors seen in humans when they experience pain and distress—facial contortions, moaning, yelping or other forms of vocalization, motor activity such as writhing, avoidance behaviors at the prospect of a repetition of the painful stimulus—can be observed in all mammals and in many other species. Likewise for the physiological signals that attend pain—activation of the sympathetic autonomous nervous system resulting in change in blood pressure, dilated pupils, sweating, increased heart rate, release of stress hormones, and so on. The discovery of cortical pain responses in premature babies shows the fallacy of relying on language as the sole criterion for consciousness (Slater et al., 2006).

**Evolutionary continuity** The first true mammals appeared at the end of the Triassic period, about 220 million years ago, with primates proliferating following the Cretaceous-Tertiary extinction event, about 60 million years ago, while humans and macaque monkeys did not diverge until 30 million years ago (Allman, 1999). *Homo sapiens* is part of an evolutionary continuum with its implied structural and behavioral continuity, rather than an independently developed organism.
While certain aspects of consciousness, in particular those relating to the recursive notion of self and to abstract, culturally transmitted knowledge, are not widespread in nonhuman animals, there is little reason to doubt that nonhuman mammals share conscious feelings—sentience—with humans. To believe that people are special, are singled out by the gift of consciousness above all other species, is a remnant of humanity’s atavistic, deeply held belief that Homo sapiens occupies a privileged place in the universe, a belief with no empirical basis.

The extent to which nonmammalian vertebrates, such as tuna, cichlid, and other fish; crows, ravens, magpies, parrots, and other birds; or even invertebrates such as the octopus or bees, with complex, nonstereotyped behaviors including delayed matching, nonmatching to sample, and other forms of learning (Giurfa, Zhang, Jenett, Menzel, & Srinivasan, 2001) are conscious is difficult to answer at this point in time (Edelman, Baars, & Seth, 2005). Without a sounder understanding of the neuronal architecture necessary to support consciousness, it is unclear where in the animal kingdom to draw the Rubicon that separates species with at least some conscious percepts from those that never experience anything and that are nothing but pure automata (Griffin, 2001).

**Arousal and states of consciousness**

There are two common, but quite distinct, usages of the term consciousness, one revolving around arousal and states of consciousness (see Schiff, chapter 78, this volume) and another one around the content of consciousness and conscious states. To be conscious of anything, the brain must be in a relatively high state of arousal (sometimes also referred to as vigilance). This statement is as true of wakefulness as it is of REM sleep that is vividly, consciously experienced—though usually not remembered—in dreams. The level of brain arousal, measured by electrical or metabolic brain activity, fluctuates in a circadian manner, and is influenced by lack of sleep, drugs and alcohol, physical exertion, and so on in a predictable manner. High arousal states are always associated with some conscious state—a percept, thought, or memory—that has a specific content. We see a face, hear music, remember an incident, plan an experiment, or fantasize about sex. Indeed, it is not clear whether one can be awake without being conscious of something. Referring to such conscious states is conceptually quite distinct from referring to states of consciousness that fluctuate with different levels of arousal.

Different levels or states of consciousness are associated with different kinds of conscious experiences. The awake state in a normal functioning individual is quite different from the dreaming state (for instance, the latter has little or no self-reflection) or from the state of deep sleep. In all three cases, the basic physiology of the brain is changed, affecting the space of possible conscious experiences. Physiology is also different in altered states of consciousness, for instance, after taking psychedelic drugs when events often have a stronger emotional connotation than in normal life. Yet another state of consciousness can occur during certain meditative practices, when interoceptive perception and insight may be enhanced compared to the normal waking state.

In some obvious but difficult to rigorously define manner, the richness of conscious experience increases as an individual transitions from deep sleep to drowsiness to full wakefulness. This richness of possible conscious experience could be quantified using notions from complexity theory that incorporate both the dimensionality and the granularity of conscious experience (e.g., Tononi, 2004; see chapter 84, this volume). For example, inactivating all of visual cortex in an otherwise normal individual would significantly reduce the dimensionality of conscious experience, since no color, shape, motion, texture, or depth could be perceived or imagined. A singular exception to this progression is REM sleep where most motor activity is shut down in the atonia that is characteristic of this phase of sleep, and the person is difficult to wake up. Yet this low level of behavioral arousal goes, paradoxically, hand in hand with high metabolic and electrical brain activity and conscious, vivid states.

Clinicians speak of impaired states of consciousness as in “the comatose state,” “the persistent vegetative state” (PVS), and the “minimal conscious state” (MCS). Here, state refers to different levels of consciousness, from a total absence in the case of coma, PVS, or general anesthesia, to a fluctuating and limited form of conscious sensation in MCS, in sleep-walking, or during a complex partial epileptic seizure (Schiff, 2004, and chapter 78, this volume).

The repertoire of distinct conscious states or experiences that are accessible to a patient in MCS is presumably minimal (possibly including pain, discomfort, and sporadic sensory percepts; but see Owen et al., 2006), immeasurably smaller than the possible conscious states that can be experienced by a healthy brain. Given the absence of any accepted theory for the minimal neuronal criteria necessary for consciousness, the distinction between a PVS patient—who shows regular sleep-wave transitions and who may be able to move eyes or limbs or smile in a reflexive manner, as in the widely publicized 2005 case of Terri Schiavo in Florida—and an MCS patient who can communicate (on occasion) in a meaningful manner (for instance, by differential eye movements) and who shows some signs of consciousness is often difficult to make in a clinical setting. Functional brain imaging of patients with global disturbances of consciousness (including akinetic mutism) reveal that dysfunction in a widespread cortical network including medial and lateral...
Prefrontal cortex and parietal associative areas is associated with a global loss of consciousness (Laureys, 2005).

In contrast to diffuse cortical damage, relatively discrete bilateral injuries to midline (paramedian) subcortical structures can also cause a complete loss of consciousness. These structures are therefore part of the enabling factors that control the level of brain arousal and that are needed for any form of consciousness to occur. For an example, consider the heterogeneous collection of more than two dozen (on each side) of nuclei in the upper brain stem (pons, midbrain, and posterior hypothalamus), collectively referred to as the reticular activating system. These nuclei—three-dimensional collections of neurons with their own cytoarchitecture and neurochemical identity—release distinct neuromodulators such as acetylcholine, noradrenaline/norepinephrine, serotonin, histamine, and orexin/hypocretin. Their axons project widely throughout the brain. These neuromodulators control the excitability of thalamus and forebrain and mediate the alternation between wakefulness and sleep, as well as the general level of both behavioral and brain arousal. Acute lesions in the reticular activating system can result in loss of consciousness and coma. Another enabling factor for consciousness is the intralaminar nuclei of the thalamus (ILN). These receive input from many brain stem nuclei and from frontal cortex and project strongly to the basal ganglia and, in a more distributed manner, into layer I of much of neocortex. Comparatively small (1 cm³ or less) bilateral lesions in the ILN can completely eliminate awareness (Bogen, 1995). Thus, the ILN are necessary for somebody to be conscious at all but do not appear to be responsible for mediating specific conscious percepts. It is likely that the specific content of any one conscious sensation is mediated by neurons in cortex and their associated satellite structures, including the amygdala, thalamus, claustrum, and basal ganglia.

The neuronal correlates of consciousness

One key objective of the inchoate science of consciousness is to search for the neuronal correlates—and ultimately the causes—of consciousness. As defined by Crick and Koch (2003), the neuronal correlates of consciousness (NCC) are the minimal neuronal mechanisms jointly sufficient for any one specific conscious percept (figure 79.1).

This definition of the NCC stresses the word “minimal,” because the question of interest is which subcomponents of the brain are actually needed. For instance, it is likely that neural activity in the cerebellum does not underlie any conscious perception and thus is not part of the NCC. That is, trains of spikes in Purkinje cells (or their absence) will not induce a sensory percept, although they may ultimately affect some behaviors.

This definition does not focus on the necessary conditions for consciousness, because of the great redundancy and parallelism found in neurobiological networks. While activity in some population of neurons may underpin a percept in one case, a different population might mediate a related percept if the former population is lost or inactivated.

Every phenomenal, subjective state will have associated NCC: one for seeing a red patch, another one for seeing Grandmother, yet a third one for feeling that a particular behavior was freely caused (feeling of agency). Perturbing or inactivating the NCC for any one specific conscious experience will affect the percept or cause it to disappear. If the NCC could be induced artificially—for instance, by cortical microstimulation in a prosthetic device or during neurosurgery—the subject will experience the associated percept. What characterizes the NCC? What are the communalities between the NCC for seeing and for hearing? Will the NCC involve all pyramidal neurons in cortex at any given point in time? Or only a varying subset of long-range projection...
cells in frontal lobes that project to the sensory cortices in the back? Only layer 5 cortical cells? Neurons that fire in a rhythmic manner? Neurons that fire in a synchronous manner? These are some of the proposals that have been advanced over the years (Chalmers, 2000).

The extent to which the NCC depends on emotions, moods, and homeostatic signals is controversial. This topic is taken up in detail by Koenigs and Adolphs, chapter 82, this volume).

Quantum mechanics and consciousness

It is implicitly assumed by neurobiologists that the relevant variables giving rise to consciousness are to be found at the neuronal level, among the synaptic releases or the action potentials in one or more population of cells, rather than at the molecular level. A few scholars have proposed that macroscopic quantum behaviors underlie consciousness. Of particular interest here is entanglement, the observation that the quantum states of multiple objects, such as two coupled electrons, may be highly correlated even though they are spatially separated, violating our intuition about locality (entanglement is also the key feature of quantum mechanics hoped to be exploited in quantum computers). The role of quantum mechanics for the photons received by the eye and for the molecules of life is not controversial. But there is no evidence that any components of the nervous system—a 37°C Celsius wet and warm tissue strongly coupled to its environment—display quantum entanglement. And even if quantum entanglement were to occur inside individual cells, molecular diffusion and action potential generation and propagation, the principal mechanism for getting information into and out of neurons, would destroy superposition. At the cellular level, the interaction of neurons is governed by classical physics (Koch & Hepp, 2006).

Interaction with the world is not required for consciousness

We are usually conscious of what goes on around us, and occasionally of what goes on within our body. So it is only natural to think that consciousness may be tightly linked to the ongoing interaction we maintain with the world and the body (O’Regan & Noe, 2001). However, there are many examples to the contrary. We are conscious of our thoughts, which do not seem to correspond to anything out there; we can also imagine things that are not out there. When we do so, cortical sensory areas can be activated from the inside, though there are some differences from normal visual perception. Also, stimulus-independent consciousness is associated with its own patterns of activation within cortex and thalamus (Mason, Norton, Van Horn, Wegner, & Grafton, 2007). During dreams, we are virtually disconnected from the environment (Hobson, Pace-Schott, & Stickgold, 2000)—hardly anything of what happens around us enters consciousness, and our muscles are paralyzed (except for eye muscles and diaphragm). Nevertheless, we are vividly conscious: all that seems to matter is that the thalamocortical system continues to function more or less as in wakefulness, as shown by neuronal recording, EEG, and neuroimaging studies performed during rapid eye movement (REM) sleep, when dreams are most intense (Maquet et al., 1996).

Neurological evidence indicates that neither sensory inputs nor motor outputs are needed to generate consciousness. For instance, retinally blind people can both imagine and dream visually if they become blind after 6–7 years of age or so (Hollins, 1985; Buchel, Price, Frackowiak, & Friston, 1998). Patients with the locked-in syndrome can be almost completely paralyzed, and yet they are just as conscious as healthy subjects (Laureys, 2005) and can compose eloquent accounts of their condition (Bauby, 1997). A transient form of paralysis is one of the characteristic features of narcolepsy. Severe cataleptic attacks can last for minutes and leave the patient collapsed on the floor, utterly unable to move or to signal, but fully aware of her surroundings (Siegel, 2000). Or consider the Californian drug addicts known as the frozen addicts who acquired some of the symptoms of severe, late-stage Parkinson’s disease, fully conscious, yet unable to move or speak (Langston & Palfreman, 1995). All six had previously taken synthetic heroin tainted with MPTP, which selectively and permanently destroyed dopamine-producing neurons in their basal ganglia.

Consciousness here and now depends on what certain parts of the brain are doing, without requiring any obligatory interaction with the environment or the body. Whether the development of consciousness requires such interactions in early childhood, though, is a different matter.

Consciousness does not require self-consciousness, introspection, or language

Consciousness is usually evaluated by verbal reports. Questions about consciousness (“Did you see anything on the screen?”) are answered by “looking inside” retrospectively and reporting what one has just experienced. So it is perhaps natural to suggest that consciousness may arise through the ability to reflect on our own perceptions: our brain would form a scene of what it sees, but we would become conscious of it—experience it subjectively—only when we, as a subject of experience, watch that scene from the inside. This suggestion is often framed in a neurobiological context by assuming that patterns of activity corresponding to “unconscious” or “subconscious” percepts form in posterior regions of the cerebral cortex involved in the categorization/association of sensory stimuli. These percepts then become conscious when mainly
anterior prefrontal and cingulate regions involved in self-representations and introspection interact with posterior cortex, perhaps by reading signals through forward connections and selectively amplifying them through back connections.

There is no doubt that the brain categorizes its own patterns of activity in the sense that neurons respond mainly to the activity of other neurons, so the brain is constantly “looking at itself.” However, this process is not necessarily understood in terms of a “subject” (the front) looking at an “object” represented in sensory cortices (the back). Leaving aside the mystery of why reflecting on something should make it conscious, this scenario is made less plausible by a common observation: when we become absorbed in some engaging task—for example, watching an engrossing movie, playing a fast-paced video game, or driving a motorcycle at high speed through traffic—we are vividly conscious without reflection or introspection. Often, we become so immersed in this rapid flow of experience—for example, during a difficult climb up a rock wall—that we may lose the sense of self, the inner voice. Perhaps the habit of thinking about consciousness has distracted the scholars who write upon such matters to devalue the unreflective nature of much of experience.

A neuroimaging study by Malach and collaborators (Hasson, Nir, Levy, Fuhrmann, & Malach, 2004) suggests that activation of prefrontal regions is not necessary for the emergence of perceptual consciousness but may be needed to reflect upon it and report it to others (however, see Bar et al., 2006 for evidence of very rapid and visual object-specific activation of orbitofrontal cortex). Indeed, it appears that self-related activity is actually shut off during highly demanding sensory tasks. Lesion studies also support the notion that perceptual consciousness may not require prefrontal cortex and, by inference, the functions it performs: A man who, at the age of 21, had fallen on an iron spike that completely penetrated through both of his frontal lobes, nevertheless went on to live a stable life—marring and raising two children—in an appropriate professional and social setting. Although displaying many of the typical frontal lobe behavioral disturbances, he never complained of loss of sensory perception, nor did he show visual or other deficits (Mataró, Jurado, García-Sánchez, Barraquer, & Costa-Jussà, 2001). Another case is that of a 27-year-old woman with massive bilateral prefrontal damage of unclear etiology (Markowitsch & Kessler, 2000). While manifesting grossly deficient scores in frontal-lobe-sensitive tests, she has no abnormal perceptual abilities (that is not to say that such patients do not suffer from subtle visual deficits; Barcelo, Suwazono, & Knight, 2000).

Finally, being conscious does not require language. Humans continually affirm consciousness through speech, describing and discussing their sensory and other experiences. So it is natural to think that speech and consciousness are somehow inextricably linked. They are not. Infants and animals cannot speak, but they are conscious and can report their experiences in other ways. And, of course, there are numerous patients who lost the ability to understand or use words and yet remained conscious.

**Consciousness and attention are independent processes**

Few would dispute that the relationship between consciousness and selective, visual attention is an intimate one. When subjects pay attention to an object, they become conscious of its various attributes; when the focus of attention shifts away, the object fades from consciousness. Indeed, more than a century of research efforts have quantified the ample benefits accrued to attended and consciously perceived events (Pashler, 1998; Braun, Koch, & Davis, 2001). This intimate connection has prompted many to posit that the two processes are inextricably interwoven, if not identical (Posner, 1994; Merikle & Joordens, 1997; Chun & Wolfe, 2000; O’Regan & Noe, 2001). Others, however, going back to the 19th century, have argued that attention and consciousness are distinct phenomena, with distinct functions and neuronal mechanisms (Iwasaki, 1993; Hardcastle, 1997; Lamme, 2003; Baars, 2005; Dehaene, Changeux, Naccache, Sackur, & Sergent, 2006; Koch & Tsuchiya, 2007). Recent psychophysical and neurophysiological evidence argues in favor of a dissociation between selective attention and consciousness, suggesting that events or objects can be attended to without being consciously perceived. Conversely, an event or object can be consciously perceived in the near absence of top-down attentional processing.

**Attention Without Consciousness** Consider that subjects can attend to a location for many seconds and yet fail to see one or more attributes of an object at that location. In lateral masking (visual crowding), the orientation of a peripherally presented grating is hidden from conscious sight but remains sufficiently potent to induce an orientation-dependent aftereffect (He, Cavanagh, & Intriligator, 1996). Montaser-Kouhsari and Rajimehr (2004) showed that an aftereffect induced by an invisible illusory contour required focal attention, even though the object at the center of attention was invisible. Naccache, Blandin, and Dehaene (2002) elicited priming for invisible words (suppressed by forward and backward masking) but only if the subject was attending to the invisible prime; without attention, the same word failed to elicit priming. In another experiment, male/female nudes attracted attention when they were rendered completely invisible by continuous flash suppression (Jiang, Costello, Fang, Huang, & He, 2006). When subjects had to discriminate the location of the masked nude from the location of a masked shuffled nude, they were at chance; without the intraocular masking, the images are clearly...
visible. Functional MRI evidence confirms attentional modulation of invisible images in primary visual cortex (Bahrami, Lavie, & Rees, 2007). In conclusion, attentional selection by itself is not sufficient for consciousness to occur.

Consciousness in the Absence of Attention When one focuses intensely on one event, the world is not reduced to a tunnel, with everything outside the focus of attention gone: we are always aware of some aspects of the world surrounding us, such as its gist. Indeed, gist is immune from inattentional blindness (Mack & Rock, 1998). In the 30 ms necessary to apprehend the gist of a scene, top-down attention cannot play much of a role (because gist is a property associated with the entire image, any process that locally enhances features is going to be of limited use; Fei-Fei, Iver, Koch, & Perona, 2007).

Take the perception of a single object (say a bar) in an otherwise empty display, a nonecological but common arrangement in many animal and human experiments. Here, what function would top-down, selective attention need to perform without any competing objects nearby? Indeed, the most popular neuronal model of attention, biased competition (Desimone & Duncan, 1995), predicts that in the absence of competition no or little attentional enhancement occurs, yet we are perfectly aware of the object and its background.

In a dual-task paradigm, the subject’s attention is drawn to a demanding central task, while at the same time a secondary stimulus is flashed somewhere in the periphery. Using the identical retinal layout, the subject performs either the central task, or the peripheral task, or both simultaneously (Sperling & Dosher, 1986; Braun & Sagi, 1990; Braun & Julesz, 1998). With focal attention engaged at the center, the subject can still distinguish a natural scene containing an animal (or a vehicle) from one that does not include an animal (or a vehicle), while being unable to discriminate a red-green bisection disk from a green-red one (Li et al., 2002). Likewise, subjects can tell male from female faces or even distinguish a famous from a nonfamous face (Reddy, Wilken, & Koch, 2004; Reddy, Reddy, & Koch, 2006), but are frustrated by computationally much simpler tasks (e.g., discriminating a rotated letter L from a rotated T). Thus, although we cannot be sure that observers do not deploy some limited amount of top-down attention in dual-task experiments that require training and concentration (that is, high arousal), it remains true that subjects can perform certain discriminations but not others in the near absence of top-down attention. And they are not guessing. They can be quite confident of their choices and “see,” albeit often indistinctly, what they can discriminate.

The existence of such dissociations—attention without consciousness and consciousness without attention—should not be surprising when considering their different functions. Attention is the set of mechanisms whereby the brain selects a subset of the incoming sensory information for higher-level processing, while the nonattended portion of the input is analyzed at a lower bandwidth, that is, with fewer processing resources. In primates, about one million fibers leave each eye and carry on the order of one megabyte per second of raw information. One way to deal with this deluge of data is to select a small fraction and process this reduced input in real time while nonattended stimuli suffer from benign neglect. By contrast, consciousness appears to be involved in providing a kind of “executive summary” of the current situation that is useful for decision making, planning, and learning (Koch, 2004; Baars, 2005).

The Neuronal Basis of Perceptual Illusions

The possibility of precisely manipulating visual percepts in time and space has made vision a preferred modality for seeking the NCC. Psychologists have perfected a number of techniques—masking, binocular rivalry, continuous flash suppression, motion-induced blindness, change blindness, inattentional blindness—in which the seemingly simple and unambiguous relationship between a physical stimulus in the world and its associated percept in the privacy of the subject’s mind is disrupted (Kim & Blake, 2005; see also Rees, chapter 80 in this volume, and Macknik & Martinez-Conde, chapter 81). A stimulus can be perceptually suppressed for minutes at a time: the image is projected into one of the observer’s eyes, but it is invisible, not seen. In this manner the neural mechanisms that respond to the subjective percept rather than the physical stimulus can be isolated, permitting the footprints of visual consciousness to be tracked in the brain.

A popular illusion is binocular rivalry (Blake & Logothetis, 2002). Here, a small image (e.g., a horizontal grating) is presented to the left eye and another image (e.g., a vertical grating) is shown to the corresponding location in the right eye. In spite of the constant visual stimulus, observers consciously see the horizontal grating alternate every few seconds with the vertical one. The brain does not allow for the simultaneous perception of both images.

Macaque monkeys can be trained to report whether they see one or the other image. The distribution of switching times and the way in which changing the contrast of one image affects the reports leave little doubt that monkeys and humans experience the same basic phenomenon. In a series of elegant experiments, Logothetis and colleagues (Leopold & Logothetis, 1996; Logothetis, 1998) recorded from a variety of visual cortical areas in the awake macaque monkey while the animal performed a binocular rivalry task. In primary visual cortex (V1), only a small fraction of cells weakly modulate their response as a function of the percept of the monkey. The majority of cells responded to one or the
other retinal stimulus with little regard to what the animal perceived at the time. In contrast, in a high-level cortical area such as the inferior temporal (IT) cortex along the ventral pathway, almost all neurons responded only to the perceptual dominant stimulus, that is, to the stimulus that was being reported. For example, when a face and a more abstract design were presented, one of these to each eye, a “face” cell fired only when the animal indicated by its performance that it saw the face and not the design presented to the other eye (figure 79.2). This result implies that the NCC involve activity in neurons in inferior temporal cortex.

Clearly this does not imply that the NCC are local to IT. Given known anatomical connections, it is likely that specific reciprocal interactions between IT cells and neurons in parts of the prefrontal cortex are necessary for the NCC. This possibility is compatible with the widely accepted notion that the NCC involve positive feedback to ensure that neural activity is persistent and strong enough to exceed some threshold and to be distributed to multiple cognitive systems, including working memory, planning, and language.

In a related perceptual phenomenon, flash suppression, the percept associated with an image projected into one eye is transiently suppressed by flashing another image into the other eye (while the original image remains; Wolfe, 1984). Its methodological advantage over binocular rivalry is that the timing of the perceptual transition is determined by an external trigger rather than by an internal event. The majority of responsive cells in inferior temporal cortex and in the superior temporal sulcus follow the monkey’s behavior—and therefore its percept (Sheinberg & Logothetis, 1997). That is, when the animal perceives a cell’s preferred stimulus, the neuron fires; when the stimulus is present on the retina but is perceptually suppressed, the cell falls silent, even though legions of V1 neurons fire vigorously to the same stimulus. Single-neuron recordings in the medial temporal lobe of epileptic patients during flash suppression likewise demonstrate abolition of their responses when their preferred stimulus is present on the retina but not seen (Kreiman, Fried, & Koch, 2002).

A related question is the extent to which a specialized network of neurons in any one cortical region mediates the NCC for all columnar properties associated with that region. This has been directly tested by recording from individual neurons in the middle temporal cortex (MT) of monkeys viewing perceptual rivalrous motion stimuli (Maier, Logothetis, & Leopold, 2007). Contrary to expectations, small changes in the stimulus configuration lead to large changes in the firing activity of cells that carry perceptual rather than purely sensory signals. Depending on which one of four stimulus configurations the physiologists used, between 70% and 90% of all MT cells can carry NCC-related signals. This result implies either that specialized cells expressing the NCC are located beyond area MT or that such specialized cells do not exist in large numbers and that almost any neuron can participate in mediating perceptual consciousness.

A number of fMRI experiments have exploited binocular rivalry and related illusions to identify the hemodynamic activity underlying visual consciousness in humans. They demonstrate quite conclusively that BOLD activity in the upper stages of the ventral pathway (e.g., the fusiform face area and the parahippocampal place area) follow the percept.

![Figure 79.2](image_url)
and not simply the retinal stimulus (Rees & Frith, 2007; Rees, chapter 80, this volume).

There is a lively debate about the extent to which neurons in primary visual cortex are directly responsible for expressing the subject’s conscious percept. That is, is V1 part of the NCC (Crick & Koch, 1995)? It is clear that retinal neurons are not part of the NCC for visual experiences. While retinal neurons often correlate with visual experience, the spiking activity of retinal ganglion cells does not accord with visual experience (for example, there are no photoreceptors at the blind spot; yet no hole in the field of view is apparent; in dreams, vivid imagery occurs despite closed eyes; and so on).

A number of compelling observations link perception with fMRI BOLD activity in human V1 and even LGN (Tong, Nakayama, Vaughan, & Kanwisher, 1998; Lee, Blake, & Heeger, 2005). These data appear to be at odds with single-neuron recordings from the monkey (but see Maier et al., 2008). It is known that modulatory, feedback signals—for example those mediating selective attention—can be much more easily detected by means of fMRI than by single-unit recordings (Wilke, Logothetis, & Leopold, 2006; Logothetis, 2008). Indeed, unless attentional effects are carefully controlled for, their neural correlates cannot be untangled from those of consciousness (Huk, Rees, & Heeger, 2001; Tse, Martinez-Conde, Schlegel, & Macknik, 2005). This aim has now been achieved in an elegant study of Lee, Blake, and Heeger (2007). Using a dual-task paradigm they find that hemodynamic BOLD activity in human V1 reflects attentional processes but does not directly correlate with the conscious percept of the subject. Haynes and Rees (2005) exploited multivariate decoding techniques to read out perceptually suppressed information (the orientation of a masked stimulus) from V1 BOLD activity, even though the stimulus orientation was so efficiently masked that subjects performed at chance levels when guessing the orientation. This finding supports the hypothesis that information present in V1 is accessible neither to behavior nor to consciousness, as hypothesized by Crick and Koch (1995).

In a powerful combination of binocular rivalry and flash suppression, a stationary image in one eye can be suppressed for minutes on end by continuously flashing different images into the other eye (continuous flash suppression; Tsuchiya & Koch, 2005; Tsuchiya, Koch, Gilroy, & Blake, 2006). This paradigm lends itself naturally to further investigate the relationship between neural activity—whether assayed at the single-neuron or at the brain-voxel level—and conscious perception (Jiang & He, 2006).

**Other questions related to perceptual consciousness**

The attributes of even simple percepts seem to vary along a continuum. For instance, a patch of color has a brightness and a hue that are variable, just as a simple tone has an associated loudness and pitch. However, is it possible that each particular, consciously experienced, percept is all-or-none? Might a pure tone of a particular pitch and loudness be experienced as an atom of perception, either heard or not, rather than gradually emerging from the noisy background? The perception of the world around us would then be a superposition of many elementary, binary percepts (Sergent & Dehaene, 2004).

Is perception continuous, like a river, or does it consist of a series of discontinuous batches, rather like the discrete frames in a movie (Purves, Paydarfar, & Andrews, 1996; VanRullen & Koch, 2003)? In cinematographic vision (Sacks, 2004), a rare form of visual migraine, the subject sees the movement of objects as fractured in time, as a succession of different configurations and positions, without any movement in between. The hypothesis that visual perception is quantized in discrete batches of variable duration, most often related to EEG rhythms in various frequency ranges (from theta to beta), is an old one. This idea is being revisited in light of the discrepancies of timing of perceptual events within and across different sensory modalities. For instance, even though a change in the color of an object occurs simultaneously with a change in its direction of motion, it may not be perceived that way (Zeki, 1998; Bartels & Zeki, 2006; Stetson, Cui, Montague, & Eagleman, 2006).

**Forward versus feedback projections**

Many actions in response to sensory inputs are rapid, transient, stereotyped, and unconscious (Milner & Goodale, 1995). They can be thought of as cortical reflexes and are sometimes called zombie behaviors (Koch & Crick, 2001). A slower, all-purpose conscious mode deals with broader, less stereotyped, and more complex aspects of the sensory input (or a reflection of these, as in imagery) and takes time to decide on appropriate responses. A consciousness mode is needed because otherwise a vast number of different zombie modes would be required to react to unusual events. The conscious system may interfere somewhat with the concurrent zombie systems (Beilock, Carr, MacMahon, & Starkes, 2002): focusing consciousness onto the smooth execution of a complex, multi-component, and highly trained sensorimotor task—dribbling a soccer ball, to give one example—can interfere with its smooth execution, something well known to athletes and their trainers. Having both a zombie mode that responds in a well-rehearsed and stereotyped manner and a slower system that allows time for planning more complex behavior is a great evolutionary discovery. This latter aspect, planning, may be one of the principal functions of consciousness.

It seems possible that visual zombie modes in the cortex mainly use the dorsal stream in the parietal region (Milner...
& Goodale, 1995). However, parietal activity can affect consciousness by producing attentional effects on the ventral stream, at least under some circumstances. The basis of this inference is clinical case studies and fMRI experiments in normal subjects (Corbetta & Shulman, 2002). The conscious mode for vision depends largely on the ventral “what” stream (but see Bar et al., 2006). Seemingly complex visual processing (such as detecting animals in natural, cluttered images) can be accomplished by cortex within 130–150 ms (Thorpe, Fize, & Marlot, 1996; VanRullen & Koch, 2003), too fast for consciousness to occur. It is plausible that such behaviors are mediated by a purely feedforward moving wave of spiking activity that passes from the retina through V1, into V4, IT, and prefrontal cortex, until it affects motor neurons in the spinal cord that control the finger press (as in a typical laboratory experiment). The hypothesis that the basic processing of information is feedforward is supported most directly by the short times required for a selective response to appear in IT cells (Perrett, Hietanen, Oram, & Benson, 1992). Indeed, Hung and colleagues (2005) were able to decode from the spiking activity 100 ms after image onset from a couple of hundred neurons in monkey IT the identity of a single image flashed onto the retina of the fixating animal. Coupled with a suitable motor output, such a feedforward network implements a zombie behavior—rapidly and efficiently subserving a binary categorization task in the absence of any conscious experience.

Conscious perception is believed to require more sustained, reverberatory neural activity, most likely by way of cortico-cortical feedback from other neocortical regions (see Macknik & Martinez-Conde, chapter 81, this volume). These feedback loops would explain why in backward masking a second stimulus, flashed 80–100 ms after onset of a first image, can still interfere (mask) with the percept of the first image. The reverberatory activity builds up over time until it exceeds a critical threshold. At this point, the sustained neural activity rapidly propagates to parietal, prefrontal, and anterior cingulate cortical regions, thalamus, claustrum (Crick & Koch, 2005), and related structures that support short-term memory, multimodality integration, planning, speech, and other processes intimately related to consciousness. Competition prevents more than one or a very small number of percepts to be simultaneously and actively represented. This is the hypothesis at the heart of the global workspace model of consciousness (Baars, 1988; Dehaene, Sergent, & Changeux, 2003). Sending visual information to more frontal structures would allow the associated visual events to be decoded and placed into context (for instance, by accessing various memory banks) and to have this interpretation feed back to the stimulus representation in visual cortex (Jazayeri & Movshon, 2007).

Conclusion

Ever since the Greeks first considered the mind-body problem more than two millennia ago, it has been the domain of armchair speculations and esoteric debates with no apparent resolution. Yet many aspects of this ancient set of questions now fall squarely within the domain of science.

It is known that consciousness does not require sensory input or motor output. Based on clinical and brain-imaging evidence, consciousness does not require self-consciousness, reflection, introspection, or language, although all these capabilities deeply enrich consciousness. Psychophysical and imaging evidence demonstrates that consciousness and selective attention can be dissociated. It appears that the neuronal correlates of consciousness require extensive but selective activity in the thalamocortical system, supported by enabling systems in the central thalamus, midbrain, and brain stem.

To further progress, it is imperative to record from a large number of neurons simultaneously at many locations throughout the thalamocortical system and related satellites (in particular the claustrum; Crick & Koch, 2005) in behaving subjects. This effort also demands a battery of behaviors (akin to but different from the well-known Turing test for intelligence; Koch & Tononi, 2008) that the subject—a newborn infant, immobilized patient, or nonhuman animal—has to pass before considering him, her, or it to possess some measure of consciousness. This is not an insurmountable step for mammals such as the monkey or the mouse that share many behaviors and brain structures with humans. For example, one particular mouse model of contingency awareness (C. Han et al., 2003) is based on the differential requirement for awareness of trace versus delay associative eyeblink conditioning in humans (Clark & Squire, 1998).

The growing ability of neuroscientists to manipulate in a reversible, transient, deliberate, and delicate manner identified populations of neurons using methods from molecular biology (Aravanis, Wang, Meltzer, & Mogri, 2007; X. Han & Boyd, 2007; Zhang, Wang, Adamantidis, de Lecea, & Deisseroth, 2007) opens the possibility of moving from correlation—observing that a particular conscious state is associated with some neural or hemodynamic activity—to causation. Exploiting these increasingly powerful tools depends on the simultaneous development of appropriate behavioral assays and model organisms amenable to large-scale genomic analysis and manipulation, particularly in mice (Lein et al., 2007).

Finally, as mentioned previously, it is not known to what extent animals whose nervous systems have an architecture considerably different from the mammalian neocortex are conscious. Furthermore, whether artificial systems, such as
computers, robots, or the World Wide Web as a whole, which behave with considerable intelligence, are or can become conscious remains speculative (Koch & Tononi, 2008). What is needed is a theory of consciousness that explains in quantitative terms what type of systems, with what architecture, can possess conscious states.

Information theory may be such a theoretical approach that establishes at the fundamental level what consciousness is, how it can be measured, and what requisites a physical system must satisfy in order to generate it (Chalmers, 1996; Tononi & Edelman, 1998).

The most promising candidate for such a theoretical framework is the integrated information theory of consciousness discussed in more detail by Tononi and Balduzzi (chapter 84, this volume).

It is the combination of fine-grained neuronal analysis in animals, with ever more sensitive psychophysical and brain-imaging techniques in patients and healthy individuals, and the development of a robust theoretical framework that lend hope that we can ultimately understand one of the central mysteries of life.

REFERENCES


