

The Neuroscience 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. These first-hand, subjective experiences pose a daunting challenge to the scientific method that, in many other areas, has proven so immensely fruitful. Science can describe events microseconds following the Big Bang, offer an increasingly detailed account of matter and how to manipulate it, and uncover the biophysical and neurophysiological nuts and bolts of the brain and its pathologies. However, this same method has as yet failed to provide a satisfactory account of how first-hand, subjective experience fits into the objective, physical universe. The brute fact of consciousness comes as a total surprise; it does not appear to follow from any phenomena in traditional physics or biology. Indeed, some modern philosophers even argue that consciousness is not logically supervenient to physics (Chalmers, 1996). Supervenience is used to describe the relationship between higher-level and lower-level properties such that the property X supervenes on property Y if Y determines X. This implies, for example, that changing Y will, of necessity, change X. In that sense, biology is supervenient to physics. Put differently, two systems that are physically alike will also be biologically alike. Yet it is not at all clear whether two physically identical brains will have the same conscious state.

Note that it is not yet generally accepted that consciousness is an appropriate subject of scientific

inquiry. A number of neuroscience textbooks provide extended details about brains over hundreds of pages yet leave out what it feels like to be the owner of such an awake brain, a remarkable omission.

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 their own theory. Nothing could be further from the truth. There is an immense amount of relevant psychological, clinical, and neuroscientific data and observations that needs to be accounted for. Furthermore, the modern focus on the neuronal basis of consciousness in the brain—rather than on interminable philosophical debates—has given brain scientists tools to greatly increase our knowledge of the conscious mind.

Consciousness is a state-dependent property of certain types of complex, biological, adaptive, and highly interconnected systems. The best example of consciousness is found in a healthy and attentive human brain, for example, the reader of this chapter. In deep sleep, consciousness ceases. Small lesions in the midbrain and thalamus can lead to a complete loss of consciousness, and 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 exploiting a number of empirical approaches that shed light on the neural basis of consciousness. This chapter reviews these approaches and summarizes what has been learned.

What Phenomena Does Consciousness Encompass?

There are many definitions of consciousness (Koch and Tononi, 2007). A common philosophical one 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 be me, to see red, or to be angry” also emphasizes the subjective or *first-person perspective* of consciousness: it is a subject, an “I,” who is having the experiences and the experience is inevitably private.

What it feels like to have a particular experience is called the *qualia* 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.

A science of consciousness must explain the exact relationship between phenomenal, mental states, and brain states. This is the heart of the classical mind-body problem: *What is the nature of the relationship between the immaterial, conscious mind and its physical basis in the electro-chemical interactions in the body?* This problem can be divided into several subproblems:

1. Why is there any experience at all? Or, put differently, why does a brain state feel like anything? In philosophy, this is referred to by some as the *Hard Problem* (note the capitalization), or as the explanatory gap between the material, objective world and the subjective, phenomenal world (Chalmers, 1996). 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, as 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, it is best to put this question aside for the moment and not be taken in by defeatist arguments.
2. Why is the relationship among different experiences the way it is? For instance, red, yellow, green, cyan, blue, 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.

3. Why are feelings private? As expressed by poets and novelists, we cannot communicate an experience to somebody else except by way of example.
4. 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.
5. Why are only some behaviors associated with conscious states? Much brain activity and associated behavior occur without any conscious sensation.

The Neurobiology of Free Will

A further aspect of the mind-body problem is the question of free will, a vast topic. Answering this question goes to the heart of the way people think of themselves. The spectrum of views ranges from the traditional and deeply embedded belief that we are free, autonomous, and conscious actors to the view that we are biological machines driven by needs and desires beyond conscious access and without willful control.

Of great relevance are the classical findings by Libet and colleagues (1983) of brain events that precede the conscious initiation of a voluntary action. In this elegant experiment, subjects were sitting in front of an oscilloscope, tracking a spot of light moving every 2.56 sec around a circle. Every now and then, “spontaneously,” subjects had to carry out a specific voluntary action, here flexing their wrist. If this action is repeated sufficiently often while electrical activity around the vertex of the head is recorded, a *readiness potential* (*Bereitschaftspotential*) in the form of a sustained scalp negativity develops long before the muscle starts to move. Libet asked subjects to silently note the position of the spot of light when they first “felt the urge” to flex their wrist and to report this location afterward. This temporal marker for the awareness of willing an action occurs on average 200 msec before initiation of muscular action (with a standard error of about 20 msec), in accordance with commonsense notions of the causal action of free will. However, the readiness potential can be detected at least 350 msec before awareness of the action. In other words, the subject’s brain signals the action at least half a second before the subject feels that he or she has initiated it!

This simple result has been replicated but, because of its counterintuitive implication that conscious will has no causal role, continues to be vigorously debated (Haggard and Eimer, 1999).

Psychological work in both normal individuals as well as in patients reveals further 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).

Yet 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, what psychologists call the *feeling of agency* or *authorship* (e.g., “I decided to lift my finger”), is certainly a subjective state with an associated quale no different in kind from the quale of a toothache or seeing marine blue. So even if free will is a complete chimera, the subjective feeling of willing an action must have some neuronal correlate.

Direct electrical brain stimulation during neurosurgery as well as fMRI experiments implicate medial premotor and anterior cingulate cortices in generating the subjective feeling of triggering an action (Lau *et al.*, 2004). In other words, the neural correlate for the feeling of apparent causation involves activity in these regions.

Consciousness in Other Species

Data about subjective states come not only from people that can talk about their subjective experiences but also from nonlinguistic competent individuals—newborn babies or patients with complete paralysis of nearly all voluntary muscles (locked-in syndrome)—and, most importantly, from animals other than humans. There are 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, monkey, humans, and whales. In particular, the macaque monkey is a powerful model organism to study visual perception because it shares with the human visual system three distinct cone photopigments, binocular stereoscopic vision, a foveated retina and similar eye movements.
- **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 sole criteria 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, whereas 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.

Although 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 other mammals share conscious feelings—sentience—with humans. To believe that humans 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 squids, or bees, with complex, non-stereotyped behaviors including delayed-matching, nonmatching-to-sample and other forms of learning (Giurfa *et al.*, 2001) are conscious is difficult to answer at this point in time. 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* 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 is as true of wakefulness as it is of REM sleep that is vividly, consciously experienced—though usually not remembered—in dreams (see Chapter 42). 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 unlikely that 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. Arousal can be measured behaviorally by the signal amplitude that triggers some criterion reaction (for instance, the sound level necessary to evoke an eye movement or a head turn toward the sound source).

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 conscious 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 as well as the granularity of conscious experience (e.g., Tononi, 2004). 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. As behavioral arousal increases, so does the range and complexity of behaviors of which an individual is capable. 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.

These observations suggest a two-dimensional graph (Fig. 53.1) in which the richness of conscious experience (its representational capacity) is plotted as a function of levels of behavioral arousal or responsiveness.

Global disorders of consciousness can likewise be mapped onto this plane. 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 (hopefully) general anesthesia, to a fluctuating and limited form of conscious sensation in MCS, sleep walking or during a complex partial epileptic seizure (Schiff, 2004).

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), immeasurably smaller than the possible conscious states that can be experienced by a healthy brain. In the limit of brain death, the origin of this space has been reached with no experience at all (Fig. 53.1). More relevant to clinical practice is the case of global anesthesia, during which the patient should not experience anything so as to avoid traumatic memories and their undesirable sequelae.

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 a 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 in a clinical setting. Functional brain imaging may prove immensely useful here.

Blood-oxygen-level-dependent (BOLD) functional magnetic resonance imaging (fMRI) demonstrated that a patient in a vegetative state following a severe traumatic brain injury exhibited the same pattern of brain activity as normal individuals when asked to imagine playing tennis or to imagine visiting all the rooms in her house (Owen *et al.*, 2006). Differential brain imaging of patients with such global disturbances of consciousness (including akinetic mutism) reveal that dysfunction in a widespread cortical network including medial

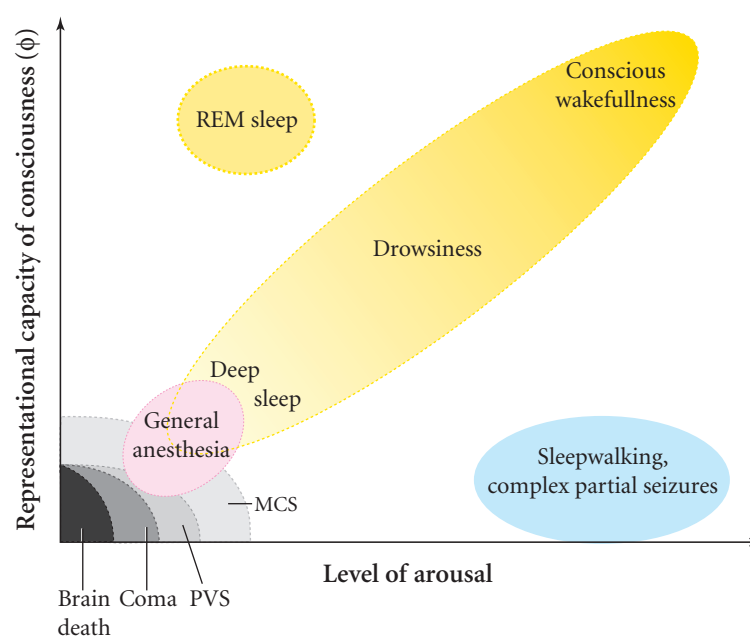


FIGURE 53.1 Normal and pathological brain states can be situated in a two-dimensional graph. Here increasing levels of behaviorally-determined arousal is plotted on the x-axis and the “richness” or “representational capacity of consciousness” is plotted on the y-axis. Increasing arousal can be measured by the threshold to obtain some specific behavior (for instance, spatial orientation to a sound). Healthy subjects cycle during a 24-hour period from deep sleep with low arousal and very little conscious experience to increasing levels of arousal and conscious sensation. In REM sleep, low levels of behavioral arousal go hand-in-hand with vivid consciousness. Conversely, various pathologies of clinical relevance are associated with little to no conscious content. Modified from Laureys (2005).

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) sub-cortical 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 (as determined by metabolic or electrical activity) 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 in the 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, nor-adrenaline/norepinephrine, serotonin, histamine, and orexin/hypocretin. Their axons project widely throughout the brain (Fig. 53.2). These neuromodulators control the excitability of thalamus and forebrain, 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. However, eventually the excitability of thala-

mus and forebrain can recover and consciousness can return (Villablanca, 2004).

Other enabling factors for consciousness are the five or more *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 an organism to be conscious at all but do not appear to be responsible for mediating specific conscious percepts or memories.

If a single substance is critical for consciousness, then acetylcholine is the most likely candidate. Two major cholinergic pathways originate in the brain stem and in the basal forebrain (Fig. 53.2). Brain stem cells send an ascending projection to the thalamus, where release of acetylcholine facilitates thalamo-cortical relay cells and suppresses inhibitory interneurons. Cholinergic cells are therefore well positioned to influence all of cortex by controlling the thalamus. In contrast, cholinergic basal forebrain neurons send their axons to a much wider array of target structures. Collectively, brain stem and basal forebrain cholinergic cells innervate the thalamus, hippocampus, amygdala, and neocortex.

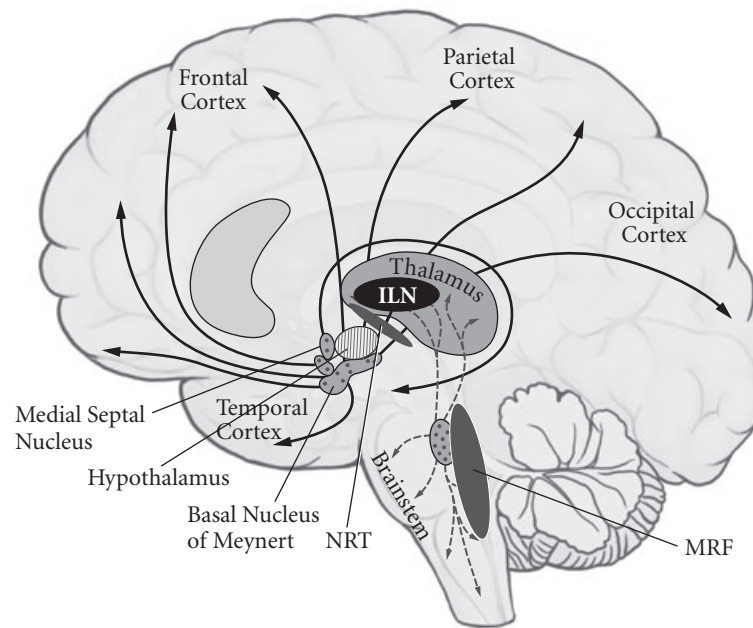


FIGURE 53.2 Midline structures in the brainstem and thalamus necessary to regulate the level of brain arousal include the intralaminar nuclei of the thalamus (ILN), the thalamic reticular nucleus (NRT) encapsulating the dorsal thalamus, and the midbrain reticular formation (MRF) that includes the reticular activating system. Small, bilateral lesions in many of these nuclei cause a global loss of consciousness.

Cholinergic activity fluctuates with the sleep–wake cycle. In general, increasing levels of spiking activity in cholinergic neurons is associated with wakefulness or REM sleep, and decreasing levels occur during non-REM or slow-wave sleep. Lastly, many neurological pathologies whose symptoms include disturbances of consciousness, such as Parkinson’s disease, Alzheimer’s disease, and other forms of dementia, are associated with a selective loss of cholinergic neurons.

In summary, a plethora of nuclei with distinct chemical signatures in the thalamus, midbrain, and pons must function for the brain of an individual to be sufficiently aroused to experience anything at all. These nuclei belong to the enabling factors for consciousness. 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 the basal ganglia.

The Neuronal Correlates of Consciousness

Progress in addressing the mind–body problem has come from focusing on empirically accessible questions rather than on eristic philosophical arguments with no clear resolution. One key objective has been 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* (Fig. 53.3).

This definition of the NCC stresses the word “minimal,” because the question of interest is which subcomponents of the brain actually are 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 (such as eye movements).

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 hearing a siren, and so on. 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 commonalities between the NCC for seeing and for

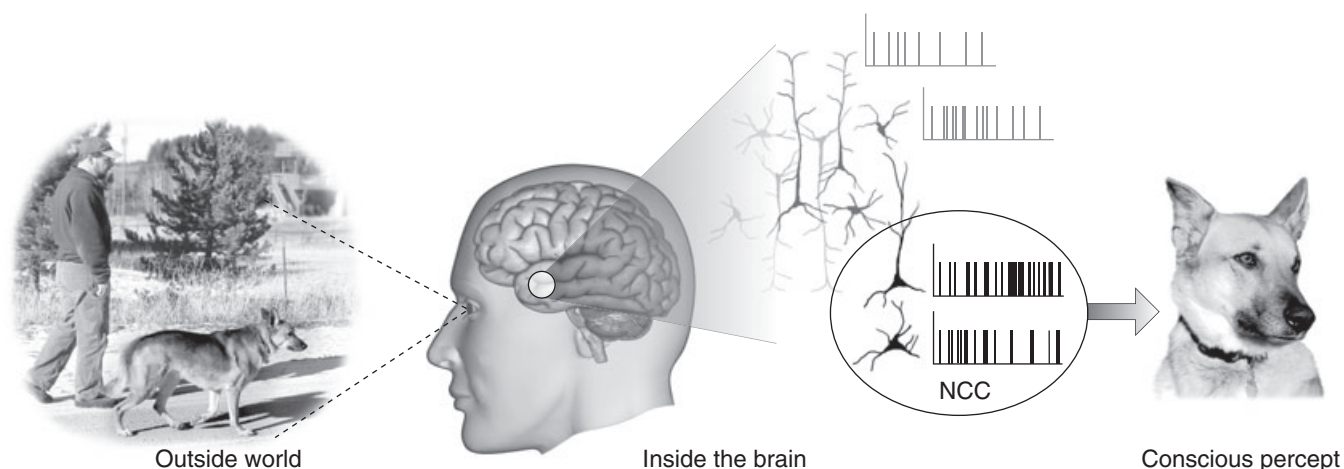


FIGURE 53.3 The Neuronal Correlates of Consciousness (NCC) are the minimal set of neural events and structures—here synchronized action potentials in neocortical pyramidal neurons—sufficient for a specific conscious percept or memory. From Koch (2004).

hearing? Will the NCC involve all pyramidal neurons in cortex at any given point in time? Or only a 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).

It is implicitly assumed by most neurobiologists that the relevant variables giving rise to consciousness are to be found at the neuronal level, among the synaptic release 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 wet and warm tissue strongly coupled to its environment—display quantum entanglement. And even if quantum entanglement were to occur inside individual cells, 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 and Hepp, 2006).

Thus, the vast majority of experiments concerned with the NCC refer to either electrophysiological

recordings in monkeys or functional brain imaging data in humans.

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, inattention 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 and Blake, 2005). With such techniques, a stimulus can be perceptually suppressed for seconds or even 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. In a *perceptual illusion*, the physical stimulus remains fixed while the percept fluctuates. The best known example is the *Necker cube*, whose 12 lines can be perceived in one of two different ways in depth (Fig. 53.4).

A perceptual illusion that can be precisely controlled is *binocular rivalry* (Blake and 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 the left or the right image. The distribution of the switching times and the way in which changing the contrast in one eye affects the reports, leaves little doubt that monkeys and humans experience the same

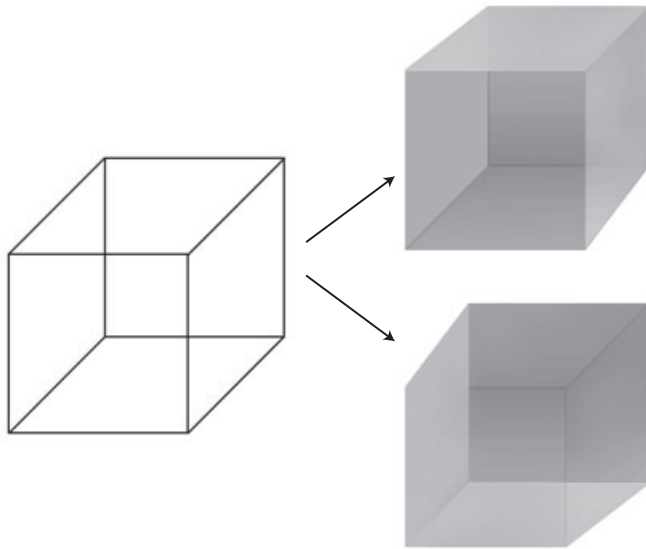


FIGURE 53.4 The left line drawing can be perceived in one of two distinct depth configurations shown on the right. Without any other cue, the visual system flips back and forth between these two interpretations of the Necker cube. From Koch (2004).

basic phenomenon. In a series of elegant experiments, Logothetis and colleagues (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 (Fig. 53.5). This result implies that the NCC involves activity in neurons in inferior temporal cortex.

Does this imply that the NCC is local to IT? At this point, this is not clear. However, given known anatomical connections, it is possible that specific reciprocal interactions between IT cells and neurons in parts of the prefrontal cortex are necessary for the NCC. This is compatible with the broadly accepted notion that the NCC must involve positive feedback to insure that neural activity is persistent and strong enough to exceed some threshold and to be broadly distributed to multiple cognitive systems, including working memory, planning, and language.

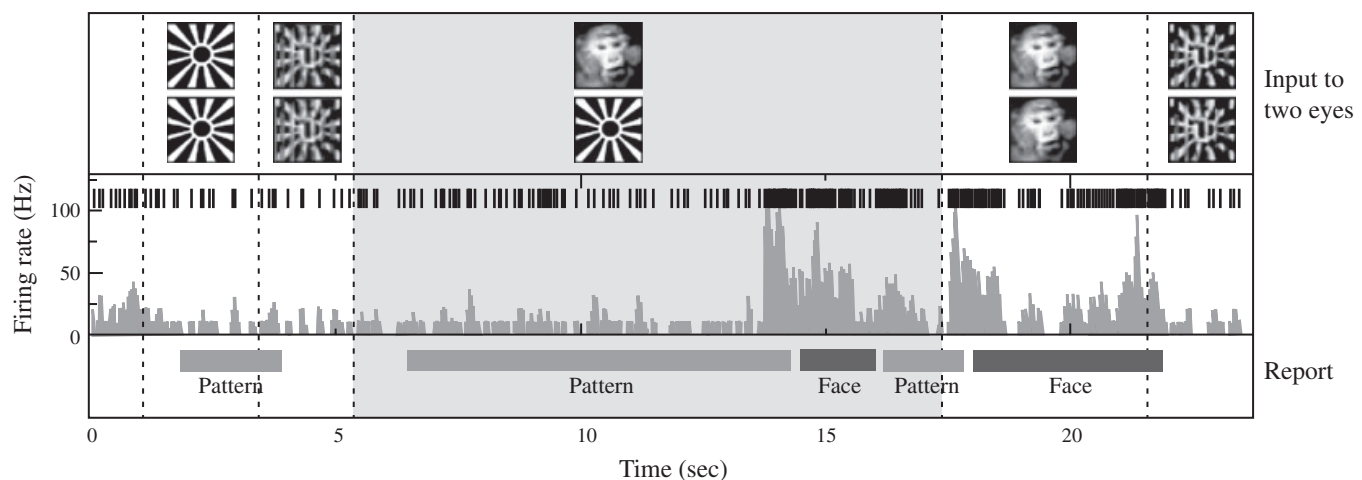


FIGURE 53.5 A fraction of a minute in the life of a typical IT cell while a monkey experiences binocular rivalry. The upper row indicates the visual input, with dotted vertical boundaries marking stimulus transitions. The second row shows the individual spikes, the third the smoothed firing rate, and the bottom row the monkey's behavior. The animal was taught to press a lever when it saw either one or the other image, but not both. The cell responded only weakly to either the sunburst design or to its optical superposition with the image of a monkey's face. During binocular rivalry (gray zone), the monkey's perception vacillated back and forth between seeing the face and seeing the bursting sun. Perception of the face was consistently accompanied (and preceded) by a strong increase in firing rate. From N. Logothetis (private communication) as modified by Koch (2004).

In a related perceptual phenomena, *flash suppression*, the percept associated with an image projected into one eye is suppressed by flashing another image into the other eye (while the original image remains). 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. 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 (Sheinberg and Logothetis, 1997). 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, and Koch, 2002).

A number of fMRI experiments have exploited binocular rivalry and related illusions to identify the hemodynamic activity underlying visual consciousness in humans. They demonstrate quite conclusively that BOLD activity in the upper stages of the ventral pathway (e.g., the fusiform face area and the parahippocampal place area) follow the percept and not simply the retinal stimulus (Rees and Frith, 2007).

There is a lively debate about the extent to which neurons in primary visual cortex simply encode the visual stimulus or are directly responsible for expressing the subject's conscious percept. That is, is V1 part of the NCC (Crick and 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 *et al.*, 1998; Lee, Blake, and Heeger, 2005). These data are at odds with single neuron recordings from the monkey. What explains this discrepancy? Most notably, the two methods operate at quite distinct spatial and temporal scales, recording with high granularity and millisecond resolution individual action potentials in monkeys versus measuring hemodynamic activity at the time scale of seconds in people. Clearly, firing rates and fMRI activity are linked but how they are linked is a very active area of research. It is, of course, always possible that the fine structure of visual cortex differs between these

two primate species. These questions have yet to be resolved.

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 trying to guess the orientation. That is, although subjects did not give any behavioral indication that they saw the orientation of the stimulus, its slant could be predicted on a single trial basis with better than chance odds from the V1 (but not from V2 or V3) BOLD signal. This finding supports the hypothesis that information present in V1 is not accessible to behavior or to consciousness.

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 and Koch, 2005). 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.

Other Perceptual Puzzles of Contemporary Interest

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 and Dehaene, 2004).

Is perception continuous, like a river, or does it consist of series of discontinuous batches, rather like the discrete frames in a movie (Purves, Paydarfar, and Andrews, 1996; VanRullen and 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 and Zeki, 2006; Stetson *et al.*, 2006).

What is the relationship between endogenous, top-down attention and consciousness? Although these are frequently coextensive—subjects are usually conscious of what they attend to—there is a considerable tradition in psychology that argues that these are distinct neurobiological processes (Koch and Tsuchiya, 2007). This question is receiving renewed attention due to the development of ever more refined and powerful visual masking techniques (Kim and Blake, 2005) that independently manipulate attention and consciousness. Indeed, it has been shown that attention can be allocated to a perceptually invisible stimulus (Naccache, Blandin, and Dehaene, 2002) and that subjects can be conscious of a stimulus without attending to it. When exploring the neural basis of these processes, it is therefore critical to not confound attention with consciousness and vice versa.

Forward versus Feedback Projections

Many actions in response to sensory inputs are rapid, transient, stereotyped, and unconscious (Milner and Goodale, 1995). They could be thought of as cortical reflexes and are characterized by rapid and somewhat stereotyped responses sometimes called *zombie behaviors* (Koch and Crick, 2002), in addition to a slower, all-purpose conscious mode. The latter, conscious mode deals more slowly 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 thoughts and 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 *et al.*, 2002): focusing consciousness onto the smooth execution of a complex, rapid, and highly trained sensory-motor 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 as well as a slower system that allows time for thinking and 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 and 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 are clinical case studies and fMRI experiments in normal subjects (Corbetta and Shulman, 2002). The conscious mode for vision depends largely on the early visual areas (beyond V1) and especially on the ventral “what” stream.

Seemingly complex visual processing (such as detecting animals in natural, cluttered images) can be accomplished by cortex within 130–150 msec (Thorpe, Fize, and Marlot, 1996; VanRullen and Koch, 2003), far too slow for conscious perception to be involved in these tasks. It is quite plausible that such behaviors are mediated by a purely feed-forward 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 feed-forward is supported most directly by the short times required for a selective response to appear in IT cells (Perrett *et al.*, 1992). Indeed, Hung *et al.* (2005) were able to decode from the spiking activity of a couple of hundred neurons in monkey IT (over intervals as short as 12.5 msec and about 100 msec after image onset), the category, identity, and even position of a single image flashed onto the retina of the fixating animal. Such hierarchical, purely feed-forward models of visual perception for object recognition in the ventral stream—an outgrowth of the original Hubel and Wiesel scheme (1968)—has been formalized and performs as well as human observers (for masked natural images) and state-of-the-art machine vision categorization algorithms (Serre *et al.*, 2007a,b). Coupled with a suitable motor output, such a feed-forward network implements a zombie behavior in the Koch and Crick (2002) sense—rapidly and efficiently subserving one task, here distinguishing animal from nonanimal pictures, in the absence of any conscious experience.

Contrariwise, conscious perception is believed to require more sustained, reverberatory neural activity, most likely via global feedback from frontal regions of neocortex back to sensory cortical areas in the back (Crick and Koch, 1995). These feedback loops would explain why in backward masking a second stimulus, flashed 80–100 msec 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, and related structures that support short-term memory, multimodality integration, planning, speech, and other processes intimately related to con-

sciousness. 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, 1989; Dehaene and Changeux, 2005). 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 sensory representation in visual cortex (Jazayeri and Movshon, 2007).

In brief, while rapid but transient neural activity in the thalamo-cortical system can mediate complex behavior without conscious sensation it is surmised that consciousness requires sustained but well-organized neural activity dependent on long-range cortico-cortical feedback.

An Information-Theoretical Theory of Consciousness

At present, 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 (as widely assumed in science fiction; e.g., the paranoid computer *HAL* in the film *2001*), remains completely speculative. What is needed is a theory of consciousness, which explains in quantitative terms what type of systems, with what architecture, can possess conscious states.

Progress in the study of the NCC on the one hand, and of the neural correlates of nonconscious zombie behaviors on the other, will hopefully lead to a better understanding of what distinguishes neural structures or processes that are associated with consciousness from those that are not. Yet such an opportunistic, data-driven approach will not lead to an understanding of why certain structures and processes have a privileged relationship with subjective experience. For example, why is it that neurons in corticothalamic circuits are essential for conscious experience, whereas cerebellar neurons, despite their huge numbers, are not? And what is wrong with cortical zombie systems that make them unsuitable to yield subjective experience? Or why is it that consciousness wanes during slow-wave sleep early in the night, despite levels of neural firing in the thalamocortical system that are comparable to the levels of firing in wakefulness?

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 and Edelman, 1998).

The most promising candidate for such a theoretical framework is the *information integration theory of consciousness* (Tononi, 2004). It posits that the most important property of consciousness is that it is extraordinarily *informative*. Any one particular conscious state rules out a huge number of alternative experiences. Classically, the reduction of uncertainty among a number of alternatives constitutes information. For example, when a subject consciously experiences reading this particular phrase, a huge number of other possible experiences are ruled out (consider all possible written phrases that could have been written in this space, in all possible fonts, ink colors, and sizes, think of the same phrases spoken aloud, or read and spoken, and so on). Thus, every experience represents one particular conscious state out of a huge repertoire of possible conscious states.

Furthermore, information associated with the occurrence of a conscious state is *integrated* information. An experience of a particular conscious state is an integrated whole. It cannot be subdivided into components that are experienced independently (Tononi and Edelman, 1998). For example, the conscious experience of this particular phrase cannot be experienced as subdivided into, say, the conscious experience of how the words look independently of the conscious experience of how they sound in the reader's mind. Similarly, visual shapes cannot be experienced independently of their color, nor can the left half of the visual field of view be experienced independently of the right half.

Based on these and other considerations, the theory claims that *a physical system can generate consciousness to the extent that it can integrate information*. This idea requires that the system has a large repertoire of available states (information) yet cannot be decomposed into a collection of causally independent subsystems (integration).

Importantly, the theory introduces a measure of a system's capacity to integrate information. This measure, called ϕ , is obtained by determining the minimum repertoire of different states that can be produced in one part of the system by perturbations of its other parts (Tononi, 2004). ϕ can loosely be thought of as the representational capacity of the system (as in Fig. 53.1). Although ϕ is not easy to calculate exactly for realistic systems, it can be estimated. Thus, by using simple computer simulations, it is possible to show that ϕ is high for neural architectures that conjoin functional specialization with functional integration, like the mammalian thalamocortical system. Conversely, ϕ is low for systems that are made up of small,

1

L1

quasi-independent modules, like the cerebellum, or for networks of randomly or uniformly connected units (Tononi, 2004).

The notion that consciousness has to do with the brain's ability to integrate information has been tested directly by transcranial magnetic stimulation. In TMS a coil is placed above the skull and a brief and intense magnetic field generates a weak electrical current in the underlying grey matter in a noninvasive manner. Massimini *et al.* (2005) compared multichannel EEG of awake and conscious subjects in response to TMS pulses to the EEG when the same subjects were deeply asleep early in the night—a time during which consciousness is much reduced. During quiet wakefulness, an initial response at the stimulation site was followed by a sequence of waves that moved to connected cortical areas several centimeters away. During slow wave sleep, by contrast, the initial response was stronger but was rapidly extinguished and did not propagate beyond the stimulation site. Thus, the fading of consciousness during certain stages of sleep may be related, as predicted by the theory, to the breakdown of information integration among specialized thalamo-cortical modules.

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.

In order to advance the resolution of these and similar questions, it will be imperative to record from a large number of neurons simultaneously at many locations throughout the cortico-thalamo system and related satellites (in particular the claustrum; Crick and Koch, 2005) in behaving subjects. Such experiments cannot, of course, be done in humans. Progress in understanding the circuitry of consciousness, therefore, demands a battery of behaviors (akin to but different from the well-known Turing test for intelligence) 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 conscious perception. 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 (Han *et al.*, 2003) is based on the differential requirement for awareness of trace versus delay associative eyeblink conditioning in humans (Clark and 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 (Han and Boyden, 2007) opens the possibility of moving from correlation—observing that a particular conscious state is associated with some neural or hemodynamic activity—to causation. For example, rather than just noting that motion perception is associated with an elevated firing rate in projection neurons in cortical area MT, we will be able to perturb the system by inactivating genetically distinct subpopulation of cells in a highly targeted manner. 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, in particularly in mice (Lein *et al.*, 2007).

It is the combination of such fine-grained neuronal analysis in mice and monkeys, ever more sensitive psychophysical and brain imaging techniques in patients and healthy individuals, and the development of a robust theoretical framework that lend hope to the belief that human ingenuity can, ultimately, understand in a rational manner one of the central mysteries of life.

References

- Bartels, A. and Zeki, S. (2006). The temporal order of binding visual attributes. *Vision Res.* **46**, 2280–2286.
- Beilock, S. L., Carr, T. H., MacMahon, C., and Starkes, J. L. (2002). When paying attention becomes counterproductive: Impact of divided versus skill-focused attention on novice and experienced performance of sensorimotor skills. *J. Exp. Psychol. Appl.* **8**, 6–16.
- Blake, R. and Logothetis, N. K. (2002). Visual competition. *Nature Reviews Neuroscience* **3**, 13–21.
- Bogen, J. E. (1995). On the neurophysiology of consciousness: I. An Overview. *Consciousness & Cognition* **4**, 52–62.
- Corbetta, M. and Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Rev. Neurosci.* **3**, 201–215.
- Chalmers, D. J. (1996). "The Conscious Mind: In Search of a Fundamental Theory." Oxford University Press, New York.
- Chalmers, D. J. (2000). What is a neural correlate of consciousness? In "Neural Correlates of Consciousness: Empirical and Conceptual Questions" (Metzinger, T., ed.) pp. 17–40. MIT Press, Cambridge.
- Crick, F. C. and Koch, C. (1995). Are we aware of neural activity in primary visual cortex? *Nature* **375**, 121–123.
- Crick, F. C. and Koch, C. (2003). A framework for consciousness. *Nature Neuroscience* **6**, 119–127.
- Edelman, D. B., Baars, J. B., and Seth, A. K. (2005). Identifying hallmarks of consciousness in non-mammalian species. *Cons. & Cogn.* **14**, 169–187.
- Giurfa, M., Zhang, S., Jenett, A., Menzel, R., and Srinivasan, M. V. (2001). The concepts of "sameness" and "difference" in an insect. *Nature* **410**, 930–933.

- Griffin, D. R. (2001). "Animal Minds: Beyond Cognition to Consciousness." University of Chicago Press, Chicago, IL.
- Haggard, P. and Eimer, M. (1999). On the relation between brain potentials and conscious awareness. *Exp. Brain Res.* **126**, 128–133.
- Han, C. J., O'Tuathaigh, C. M., van Trigt, L., Quinn, J. J., Fanselow, M. S., Mongeau, R., Koch, C., and Anderson, D. J. (2003). Trace but not delay fear conditioning requires attention and the anterior cingulate cortex. *Proc Natl. Acad. Sci. USA* **100**, 13087–13092.
- Hayes, J.-D. and Rees, G. (2005). Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nature Neurosci.* **8**, 686–691.
- Jazayeri, A. and Movshon, J. A. (2007). A new perceptual illusion reveals mechanisms of sensory decoding. *Nature* **446**, 912–915.
- Kim, C.-Y. and Blake, R. (2005). Psychophysical magic: Rendering the visible "invisible." *Trends Cogn. Sci.* **9**, 381–388.
- Koch, C. (2004). "The Quest for Consciousness: A Neurobiological Approach." Roberts, Denver, CO.
- Koch, C. and Crick, F. C. (2001). On the zombie within. *Nature* **411**, 893.
- Koch, C. and Hepp, K. (2006). Quantum mechanics and higher brain functions: Lessons from quantum computation and neurobiology. *Nature* **440**, 611–612.
- Koch, C. and Tononi, G. (2007). Consciousness. In "New Encyclopedia of Neuroscience." Elsevier, in press.
- Koch, C. and Tsuchiya, N. (2007). Attention and consciousness: Two distinct brain processes. *Trends Cog. Sci.* **11**, 16–22.
- Kreiman, G., Fried, I., and Koch, C. (2002). Single-neuron correlates of subjective vision in the human medial temporal lobe. *Proc Natl. Acad. Sci. USA* **99**, 8378–8383.
- Laureys, S. (2005). The neural correlate of (un)awareness: Lessons from the vegetative state. *Trends Cogn. Sci.* **9**, 556–559.
- Lee, S. H., Blake, R., and Heeger, D. J. (2005). Traveling waves of activity in primary visual cortex. *Nature Neurosci.* **8**, 22–23.
- Lein, E. S., Hawrylycz, M. J. et al. (2007). Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445**, 168–176.
- Libet, B., Gleason, C. A., Wright, E. W., and Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential): The unconscious initiation of a freely voluntary act. *Brain* **106**, 623–642.
- Logothetis, N. (1998). Single units and conscious vision. *Philosophical Transactions Royal Society of London B*, **353**, 1801–1818.
- Massimini, M., Ferrarelli, F., Huber, R., Esser, S. K., Singh, H., and Tononi, G. (2005). Breakdown of cortical effective connectivity during sleep. *Science* **309**, 2228–2232.
- Milner, A. D. and Goodale, M. A. (1995). "The visual brain in action." Oxford University Press, Oxford, UK.
- Owen, A. M., Cleman, M. R., Boly, M., Davis, M. H., Laureys, S., and Pickard, J. D. (2006). Detecting awareness in the vegetative state. *Science* **313**, 1402.
- Rees, G. and Frith, C. (2007). Methodologies for identifying the neural correlates of consciousness. In "The Blackwell Companion to Consciousness." (Velmans, M. and Schneider, S., eds.) pp. 553–566. Blackwell, Oxford, UK.
- Sacks, O. (2004). In the river of consciousness. *New York Review Books* **51**, 41–44.
- Schiff, N. D. (2004). The neurology of impaired consciousness: Challenges for cognitive neuroscience. In "The Cognitive Neurosciences III." (Gazzaniga, M. S., ed.) pp. 1121–1132. MIT Press, Cambridge, MA.
- Sergent, C. and Dehaene, S. (2004). Is consciousness a gradual phenomenon? Evidence for an all-or-none bifurcation during the attentional blink. *Psychol. Sci.* **15**, 720–728.
- Sheinberg, D. L. and Logothetis, N. K. (1997). The role of temporal cortical areas in perceptual organization. *Proc. Natl. Acad. Sci. USA* **94**, 3408–3413.
- Slater, R., Cantarella, A., Gallella, S., Worley, A., Boyd, S., Meek, J., and Fitzgerald, M. (2006). Cortical pain responses in human infants. *J. Neurosci.* **26**, 3662–3666.
- Stetson, C., Cui, X., Montague, P. R., and Eagleman, D. M. (2006). Motor-sensory recalibration leads to reversal of action and sensation. *Neuron* **51**, 651–659.
- Thorpe, S., Fize, D., and Marlot, C. (1996). Speed of processing in the human visual system. *Nature* **381**, 520–522.
- Tong, F., Nakayama, K., Vaughan, J. T., and Kanwisher, N. (1998). Binocular rivalry and visual awareness in human extrastriate cortex. *Neuron* **21**, 753–759.
- Tononi, G. (2004). An information integration theory of consciousness. *BMC Neuroscience* **5**, 42–72.
- Tononi, G. and Edelman, G. M. (1998). Consciousness and Complexity. *Science* **282**, 1846–1851.
- Tsuchiya, N. and Koch, C. (2005). Continuous flash suppression reduces negative afterimages. *Nature Neurosci.* **8**, 1096–1101.
- Villablanca, J. R. (2004). Counterpointing the functional role of the forebrain and of the brainstem in the control of the sleep-waking system. *J. Sleep Res.* **13**, 179–208.
- Zeki, S. (1998). Parallel processing, asynchronous perception, and a distributed system of consciousness in vision. *Neuroscientist* **4**, 365–372.

Christof Koch