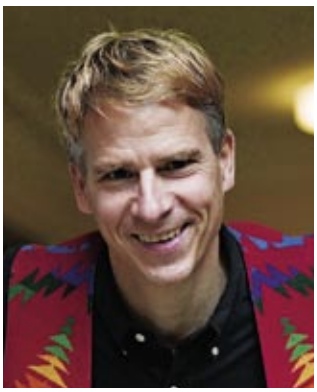
How Does Consciousness Happen?

[THE AUTHOR]



Christof Koch is professor of cognitive and behavioral biology at the California Institute of Technology, where he teaches and has conducted research on the neuronal basis of visual attention and consciousness for more than two decades. He is an avid hiker and rock climber who has scaled several noted peaks.

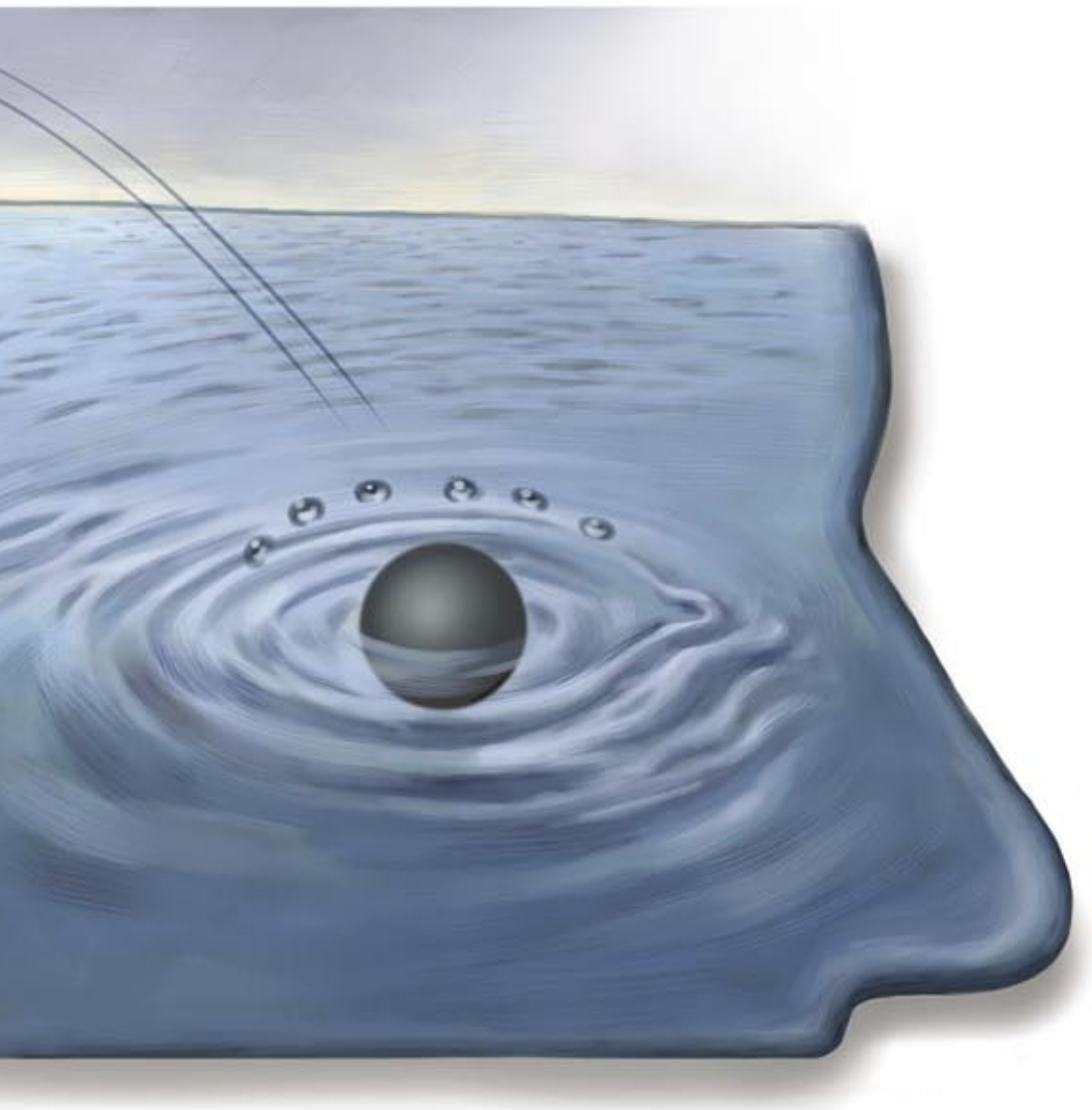
■ **HIS THEORY: For each conscious experience, a unique set of neurons in particular brain regions fires in a specific manner.**

How brain processes translate to consciousness is one of the greatest unsolved questions in science. Although the scientific method can delineate events immediately after the big bang and uncover the biochemical nuts and bolts of the brain, it has utterly failed to satisfactorily explain how subjective experience is created.

As neuroscientists, both of us have made it our life's goal to try to solve this puzzle. We share many common views, including the important acknowledgment that there is not a single problem of consciousness. Rather, numerous phenomena must be explained—in particular, self-consciousness (the ability to examine one's own desires and thoughts), the content of consciousness (what you are actually conscious *of* at any moment), and how brain processes relate to consciousness and to nonconsciousness.

So where does the solution begin? Neuroscientists do not yet understand enough about the brain's inner workings to spell out exactly how consciousness *arises* from the electrical and chemical activity of neurons. Thus, the big first

JULIA BAIER (*Koch*); MATT COLLINS (*illustration*)



Two leading neuroscientists, Christof Koch and Susan Greenfield, disagree about the activity that takes place in the brain during subjective experience

step is to determine the best neuronal correlates of consciousness (NCC)—the brain activity that matches up with specific conscious experiences. When you realize you are seeing a dog, what has happened among which neurons in your brain? When a feeling of sadness suddenly comes over you, what has happened in your brain? We are both trying to find the neuronal counterpart of each subjective experience that an individual may have. And this is where we differ.

Our disagreement over the best NCC emerged during a lively debate between us at the University of Oxford in the summer of 2006, sponsored by the Mind Science Foundation in San Antonio. Since then, we have continued to explore and challenge each other's views, a dialogue that has resulted in the article here. We are bound, nonetheless, by one fundamental commonality: our views stem primarily from neuroscience, not just philosophy. We both have considered a tremendous amount of neuroscientific, clinical and psychological data, and it is from these observations that our arguments arise.

—Christof Koch and Susan Greenfield

COURTESY OF SUSAN GREENFIELD

[THE AUTHOR]

Susan Greenfield is professor of pharmacology at the University of Oxford, director of the Royal Institution of Great Britain and member of the British Parliament's House of Lords. Her research focuses on novel brain mechanisms, including those underlying neurodegenerative diseases. Her favorite pastimes are squash and dancing.

■ **HER THEORY: For each conscious experience, neurons across the brain synchronize into coordinated assemblies, then disband.**

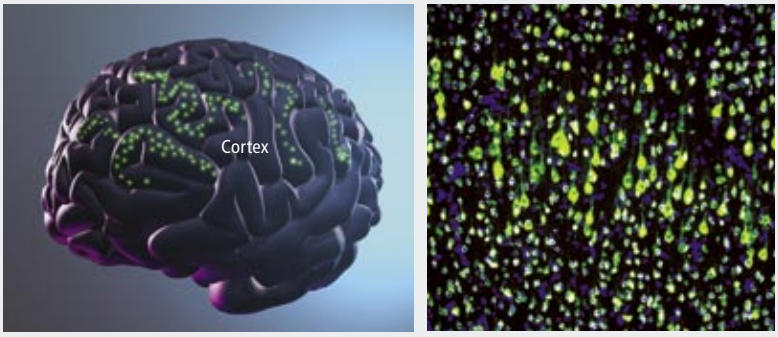
[BASIC ARGUMENTS]

CONSCIOUSNESS EXPLAINED

What happens in your brain when you see a dog, hear a voice, suddenly feel sad or have any other subjective experience?

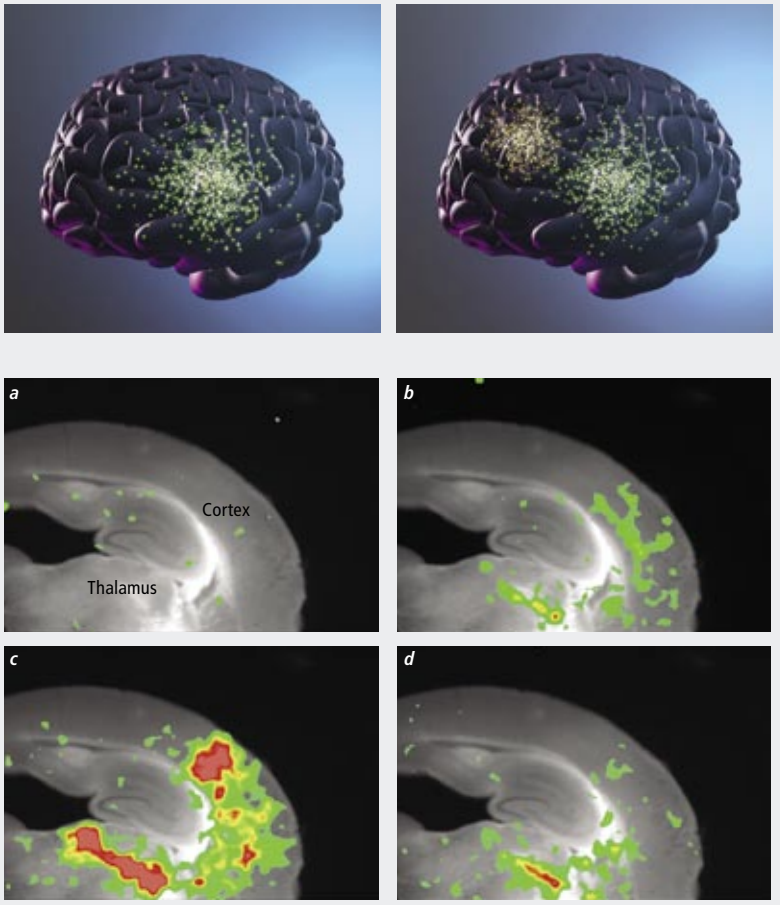
KOCH'S MODEL

A coalition of pyramidal neurons linking the back and front of the cortex fires in a unique way. Different coalitions activate to represent different stimuli from the senses (*left*). In a mouse cortex (*right*) these pyramidal cells (*green*) lie in brain layer 5, surrounded by nonneuronal cells (*blue*).



GREENFIELD'S MODEL

Neurons across the brain fire in synchrony (*green*) and prevail until a second stimulus prompts a different assembly to arise (*orange*). Various assemblies coalesce and disband moment to moment, while incorporating feedback from the body. In a rat brain (*bottom*), an assembly in the cortex forms (*a, b*), peaks (*c*), then decays (*d*) within 0.35 second after the thalamus is electrically stimulated.



Koch Speaks

"Specific groups of neurons mediate distinct conscious experiences."

Both Susan Greenfield and I are searching for the most appropriate neuronal correlates of consciousness. If we can find the right NCC, the direct cause-and-effect mechanisms that *create* consciousness may follow.

In my view, which has evolved since Francis Crick and I began investigating consciousness in 1988, every conscious percept (how the brain represents stimuli from the senses) is associated with a specific coalition of neurons acting in a specific way. There is a unique neuronal correlate of consciousness for seeing a red patch, another for seeing one's grandmother, a third for feeling angry. Perturbing or halting any neuronal correlate of consciousness will alter its associated percept or cause that percept to disappear.

Physiologically, the likely substrate for NCC is a coalition of pyramidal neurons—a type of neuron that communicates over long ranges—within the cerebral cortex. Perhaps only a million such neurons—out of the 50 billion to 100 billion in our heads—are needed to form one of these coalitions. When, say, Susan enters a crowded room and I see her face, a coalition of neurons suddenly chatters in concert for a fraction of a second or longer. The coalition reaches from the back of the cortex, where representations of visual stimuli are first processed, into the front of the cortex, which carries out executive functions such as providing perspective and enabling planning. Such a coalition would be reinforced if I paid attention to the stimulus of her image on my retina, which would strengthen the amplitude or the synchrony of the activity among the select neurons. The coalition sustains itself and suppresses competing coalitions by feeding excitatory signals back and forth among the neurons in the back and front of the cortex. If, suddenly, someone calls my name, a different coalition of neurons in the auditory cortex arises. This coalition establishes two-way communication with the front of the brain and focuses my consciousness on the voice, suppressing the earlier coalition representing Susan's face, which fades from my awareness.

One universal lesson from biology is that organisms evolve specific gadgets, and this is true for the brain. Nerve cells have developed myriad shapes and functions, along with specific wiring patterns among them. This heterogeneity is re-

STUART C. SEALFON AND POKMAN CHAN *Mount Sinai School of Medicine (mouse cortex)*; MICHAEL HILL *University of Oxford (video stills)*; ALFRED T. KAMAJIAN (*brains*)

flected in the neurons that constitute NCC. It is here that I differ most from Susan. In my view, consciousness is not some holistic property of a large collection of firing neurons that are bathed in a solution of neurotransmitters, as she argues. Instead I maintain that specific groups of neurons mediate, or even generate, distinct conscious experiences.

And soon enough, the growing ability of neuroscientists to delicately manipulate populations of neurons will move us from observing that a particular conscious state is associated with some neuronal activity to pinpointing causation—observing that a given population is partially or wholly responsible for a conscious state.

But how do we determine which set of neurons, and what activity among them, constitutes a conscious percept? Do NCC involve all the pyramidal neurons present in the cerebral cortex at any given time? Or do they just involve a subset of long-range projection cells communicating between the frontal lobes and the sensory cortices in the back of the brain? Or do they involve neurons anywhere that are firing in synchrony?

Much of the contemporary work on NCC has concentrated on vision. Visual psychologists have perfected techniques to hide things from our conscious perception, like a magician who misdirects us so that we do not see what is happening in front of our eyes. One example is flash suppression, a phenomenon discovered by then graduate student Naotsugu Tsuchiya and myself in 2005. Perception of a small, stationary image shown to one eye—say, a faint, gray, angry face projected into the right eye—is completely suppressed by a stream of constantly changing color patches flashed into the other eye. This suppression can last for minutes, even though the scary face is perfectly visible if the viewer blinks his or her left eye; although legions of neurons in the primary visual cortex are firing vigorously in response to the stimulation from the left eye, they do not contribute to consciousness. This result is hard to explain in Susan's view that any coherent firing by a large collection of neurons is a correlate of consciousness. Researchers are using such illusions to find NCC in the brains of trained monkeys and humans.

Before Francis passed away, he and I offered several proposals about how consciousness works, based on experimental results. One is that NCC include pyramidal neurons that are strategically located in an output zone of the cerebral cortex known as layer 5. These cells send out signals to, and directly receive strong excit-

[POINT/COUNTERPOINT]

BRIAN CRONIN

Why does an alarm clock induce consciousness in a sleeping (unconscious) person?

Koch's view: Neurons in a region of the brain stem called the locus coeruleus respond to a sudden, large input from the auditory nerve. They spring into action, widely broadcasting a chemical signal to the thalamus and the cerebral cortex. Other neurons release the neurotransmitter acetylcholine throughout the brain. The net effect is that the cerebral cortex and its satellite structures become aroused. Once that occurs, a widespread but tightly interconnected grouping of neurons in the auditory cortex, and its counterparts in the front of the brain and in the medial temporal lobes that support planning and memory, establishes a stable coalition using recurrent feedback. This activity takes only a fraction of a second and causes you to become conscious of the alarm.

Greenfield's view: Any strong sensory stimulus, such as a bright light, will induce consciousness, so no one particular area of the brain can be responsible for waking you up. The alarm clock prompts consciousness not because of the quality of the stimulus (in this case, auditory) but because of its quantity (loudness). Transient neuronal assemblies—many neurons acting in concert—correlate with varying degrees of consciousness: the size of an assembly from one moment to the next is determined by how readily neurons can be corralled into transient synchrony. One key factor is the strength of sensory stimulation, the effects of which are akin to a stone thrown in a pond. The larger the stone, the more extensive the ripples on the water. The louder the alarm (or brighter the light), the more likely it will be to recruit an extensive assembly of neurons, and the more extensive the assembly, the more likely that you will be awakened.

"Neuroscience needs a theory that can predict whether a fruit fly, a dog, an unresponsive Alzheimer's patient or the World Wide Web is conscious."

atory inputs from, another set of pyramidal neurons in a different region. Such an arrangement could implement a positive feedback loop, a coalition of neurons that, once triggered, would keep on firing until shut off by another coalition of neurons. These groups also fire over fractions of a second, much closer to the timescale of conscious awareness than single-neuron firings.

This notion about networks of neurons has received a boost from recent results by researchers at the Mount Sinai School of Medicine, Columbia University and the New York State Psychiatric Institute, working under Stuart C. Sealfon of Mount Sinai and Jay A. Gingrich of Columbia. Sealfon's and Gingrich's teams have demonstrated in genetically modified mice that hallucinogens—such as LSD, psilocybin (an ingredient of mushrooms) and mescaline—act on a type of molecule, called a serotonin receptor, found on the pyramidal cells that cluster in layer 5. The hypothesis that the mind-bending effects of hallucinogenic compounds come from activation of one receptor type on a specific set of neurons—rather than from "messing up" the brain's circuits in some holistic manner—can be further tested with molecular tools that can toggle layer 5 pyramidal cells on and off until the exact set of neurons being affected is identified.

A second proposal for how NCC underlie consciousness involves the claustrum, a sheet-like structure within the cortex. Remarkably the neurons composing this structure receive input from almost all regions of the cortex and project back to almost all as well. This structure may be perfectly situated to bind the activity of the sensory cortices into a single, coherent percept.

To advance these ideas, neuroscientists must sample the chattering electrical activity of a very large number of neurons at many locations. This work is delicate and difficult, but the miniaturization of electrodes is making it possible. Preliminary efforts confirm that specific groups of neurons express the types of perceptions that form our daily experiences.

None of these insights imply that one, 100 or even one million neurons living in a lab dish could be conscious. Neurons are part of vast networks and can generate consciousness only in that context. An analogy is helpful: although DNA molecules in a cell spell out the composition of the proteins in our bodies, many other molecules must also be present in the cell to construct and maintain those proteins.

The varying extent and provenance, or origin, of coalitions of neurons can also account for the different content of consciousness in infants, adults and animals. That any coalition can exist at all depends on the existence of arousal circuits in the brain stem and thalamus (which relays sensory inputs to the cortex) that are continuously active and that perfuse the cortex and its satellite structures with neurotransmitters and other substances. If a person's arousal circuits are silent—as they are when one is in deep sleep or under anesthesia or when one has suffered trauma akin to that of Terri Schiavo, the woman who fell into a persistent vegetative state that captivated the media—no stable coalition of cortical neurons can arise and the person is not conscious.

Although this model can be tested by physiological experiments, a valid criticism is that it is not a theory built from a set of principles—that is, it cannot predict what type of system has conscious experiences. Neuroscience needs a theory that predicts, based on physical measurements, which of the following organisms is conscious: a fruit fly, a dog, a human fetus five months after conception, an unresponsive Alzheimer's patient, the World Wide Web, and so on.

Some experts, including Giulio Tononi of the University of Wisconsin–Madison, are working on such theories. But we are still so ignorant about the brain that we can only speculate. Specific hypotheses that can be tested with today's technology will help. As Francis was fond of saying, what drove his and James Watson's 1953 discovery of the double-helical structure of DNA were experiments, not a theory of how genetic information might be encoded in molecules.

Fundamentally, my explanation is that qualitative, not quantitative, differences in neuronal activity give rise to consciousness. What matters is not the sheer number of neurons involved, as Susan stresses, but the informational complexity that they represent. A specific network of neurons is needed for a specific percept, not any random collection of neurons that become highly active. Furthermore, for full consciousness, a coalition of neurons must encompass both sensory representation at the back of the cortex as well as frontal structures involved in memory, planning and language. The brain works not by dint of its bulk properties but because neurons are wired up in amazingly specific and idiosyncratic patterns. These patterns reflect the accumulated information an organism has learned over its lifetime, as well as that of its ancestors, whose information is represented in genes. It is not crucial that a sufficient number of neurons are active together but that the right ones are active.

Greenfield Speaks

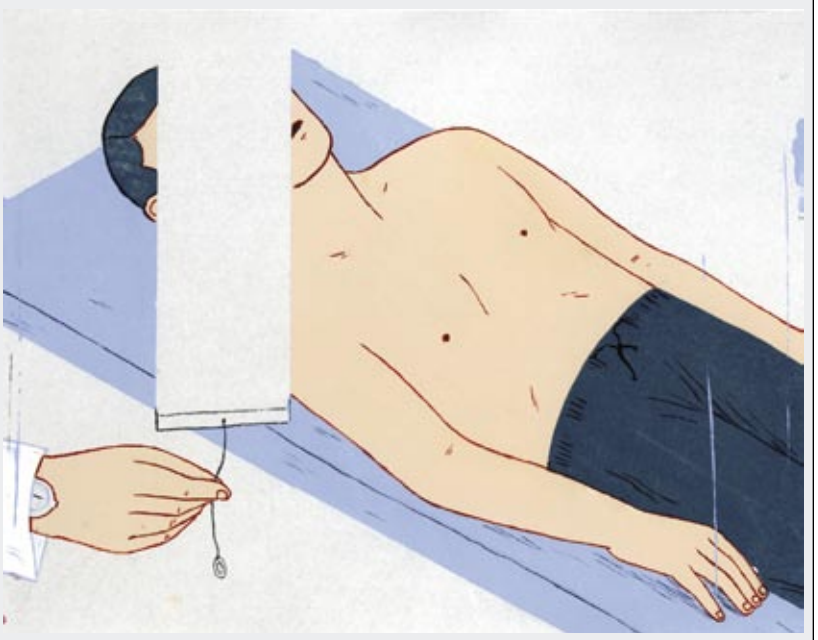
"Consciousness is generated by a quantitative increase in the holistic functioning of the brain."

If neuronal correlates of consciousness are nothing more than the discharges of certain neurons and not others, as Christof Koch suggests, then consciousness resides in the neurons themselves. Yet Christof offers no explanation as to what qualitative property such neurons or regions have, compared with others. Moreover, if not even a million neurons can generate consciousness without being part of "vast networks," then the burden of identifying NCC shifts to describing what these networks are. By looking at specific brain connections for different forms of consciousness, Christof is guilty of a 21st-century form of phrenology, in which different functions are related directly to different brain regions, especially the cortex. His enthusiasm for the cortex should be tempered by the fact that many species, such as birds, have no cerebral cortex yet are considered conscious. Even if such compartmentalization were possible, it would not explain how consciousness is generated.

In my view, consciousness cannot be divvied up into different, parallel experiences. Indeed, we know that visual stimulation can change how we hear, and vice versa. This merging of the sensorium's components argues against concepts such as an isolated visual consciousness. Most important, either you are conscious or you are not. In Christof's lab, subjects are conscious throughout experiments performed on their neurons; therefore, it is not consciousness that the experiments manipulate but the *content* of that consciousness. Any consequent explanation is really a foray into answering "What is attention?" That question is certainly valid, but it is a different one from "What is consciousness?" I contend that to define the best NCC we must elucidate the difference between consciousness and unconsciousness.

My own starting assumption is that there is no intrinsic, magical quality in any particular brain region or set of neurons that accounts for consciousness. We need to identify a special *process* within the brain. And to be a truly robust correlate of consciousness, this neuronal process must account for a variety of everyday phenomena, including the efficacy of an alarm clock, the action of anesthetics, the distinction of dreams from wakefulness, the existence of self-consciousness, the possible difference between human and animal consciousness, and the possible existence of fetal consciousness. A more plausible view of consciousness is that it is not generated by a qualitatively distinct property of the brain but by a quantitative increase in the holistic functioning of the brain. Consciousness grows as brains grow.

[POINT/COUNTERPOINT]



BRIAN CRONIN

How do anesthetics work?

Koch: Today's anesthesiologists administer a diverse collection of chemicals. Yet all abolish consciousness. Scientists used to believe that anesthetics interfered systemically with lipids in the cellular membranes of neurons. But we now know that the compounds interfere with various neuronal processes by binding to certain membrane proteins. There is no single unique mechanism that causes consciousness to stop functioning. Among the most important causes, however, is that anesthesia strengthens synaptic inhibition, or reduces synaptic excitation, in large regions of the brain. Activity is not fully shut down, but the ability of groups of neurons to form stable coalitions is severely compromised. When neurons that encompass the back and the front of the cerebral cortex cannot set up synchronized communication, consciousness becomes impossible.

Greenfield: Anesthetics do not switch off any single brain area; they depress neuronal activity in different regions across the whole brain. Anesthetics therefore achieve their effect by altering an emergent property of the holistic brain: neuronal assemblies. As anesthetics reduce the size of neuronal assemblies, they reduce the degree of consciousness until it is nonexistent. This scenario also explains the different stages of consciousness that can occur as anesthesia takes effect, such as hyperexcitability and delirium. I have suggested elsewhere that people who have brains with underfunctioning neuronal connections, and who hence have small assemblies, often exhibit strong emotions and lack of reason—just the types of states many patients exhibit while anesthesia is taking effect and their assemblies are shrinking.

[POINT/COUNTERPOINT]

Why is there a subjective difference between dreaming and wakefulness?

Koch: Although the brain is highly active during the rapid eye movement phase of sleep that is most associated with vivid dreams, the regional pattern of brain activity is quite distinct from that of wakefulness. In particular, the limbic system (loosely, the system of emotions and memory) is very active, but the parts of the frontal lobes that are involved in rational thought are subdued. In both dreaming and wakefulness, coalitions of neurons form, but they include neurons in different parts of the brain. During wakefulness, the coalitions include many more neurons in the prefrontal cortex, where reason and sensible narratives are imposed to order perceptions, but that activity is notably lacking during dreams. These features reflect the often bizarre and strong emotional content of dreaming.

Greenfield: Dreams most likely correlate with assemblies of neurons that are much smaller than those occurring when we are awake. The assemblies would be limited because no strong external stimuli are engaging large numbers of neurons. The transient recruitment of neurons during dreams is thus driven purely by response to spontaneous, intrinsic brain activity. And because the assemblies are not triggered by a sequential narrative of events in the outside world, the linkages among assemblies are haphazard, idiosyncratic or nonexistent, leaving dreams as random images or events. The lack of extensive, operational neuronal connections would also account for the notable absence of the checks and balances that normally characterize adult cognition when awake.

But what is the key neuronal mechanism in this process? The attempt to show a process-related correlate of consciousness has been inspired by various findings, including those of German neurophysiologist Wolf Singer. Singer demonstrated that a huge population of neurons between the thalamus and the cerebral cortex transiently fire together at a frequency of 40 times a second. But because the same activity can arise in this tissue kept alive in a lab dish, an additional condition must be a prerequisite for consciousness.

Neuroscientist Rodolfo Llinas of New York University Medical Center more recently suggested that this coordinated, transient firing sets up two complementary loops between the thalamus and the cerebral cortex that work in conjunction to maintain consciousness: a "specific" system relating to the content of consciousness and a "nonspecific" system relating to the arousal and alertness of consciousness. This account does indeed provide an explanation for why the strong sensory input of an alarm clock triggers full consciousness. Moreover, Llinas's model distinguishes between the consciousness of dreams and that of wakefulness; in dreams, there is no sensory input to feed the arousal loop, so only the content loop functions.

The central problem is that models developed by Llinas and others conceive of consciousness as an all-or-nothing condition. They fail to describe how the physical brain can accommodate the ebb and flow of a continuously variable conscious state. I favor an alternative. For more than a decade, scientists have known that the activity of tens of millions of neurons can synchronize for a few hundred milliseconds, then disband in less than a second. These "assemblies" of coordinating cells can vary continuously in just the right space and timescales for the here-and-now experience of consciousness. Wide-ranging networks of neurons assemble, disassemble and reassemble in coalitions that are unique to each moment. My model is that consciousness varies in degree from one moment to the next and that the number of neurons active within an assembly correlates with the degree of consciousness present at any given time.

This neuronal correlate of consciousness—the transient assembly—satisfies all the items on the shopping list of phenomena above. The efficacy of an alarm clock is explained as a very vigorous sensory input that triggers a large, synchronous assembly. Dreams and wakefulness differ because dreams result from a small assem-

BRIAN CRONIN

bly driven by weak internal stimuli, whereas wakefulness results from a larger assembly driven by stronger external stimuli. Anesthetics restrict the size of assemblies, thus inducing unconsciousness. Self-consciousness can arise only in a brain large and interconnected enough to devise extensive neuronal networks. The degree of consciousness in an animal or a human fetus depends on the sizes of their assemblies, too.

Recall that neither Christof nor I is attempting to explain *how* consciousness arises. We are not attempting to answer what Australian philosopher David Chalmers has dubbed the "hard problem": determining how physiological events in the brain translate into what *you* experience as consciousness. We are seeking a correlation—a way to show how brain phenomena and subjective experiences match up, without identifying the all-important middle step of how a phenomenon causes an experience. Neuronal assemblies do not "create" consciousness but rather are indices of degrees of consciousness. Because an assembly's size and the corresponding degree of consciousness result from a variety of physiological factors—such as degree of connectivity, strength of stimuli and competition from other assemblies—each factor could eventually be manipulated experimentally. The assembly model's ability to generate falsifiable hypotheses *and* account for the diverse range of phenomena related to consciousness surely makes it particularly powerful.

An obvious criticism of the assembly model, which Christof articulated during our Oxford debate, is that it merely posits that "size is everything." But most of science is indeed "all about measurement"—the objective quantification of observations. Size *is* everything in science. Other skeptics say that assemblies are too vague a notion, but several researchers have revealed detailed characterizations of neuronal mechanisms that underlie the generation of assemblies lasting less than a second, such as Amiram Grinvald of the Weizmann Institute of Science in Rehovot, Israel, Ole Paulsen of Oxford and John G. Jefferys of the University of Birmingham in England.

Decisive tests in humans must await better noninvasive imaging techniques that have a time resolution commensurate with the milliseconds-long timescale of the formation and disbanding of neuronal assemblies. Once these techniques are available, we should be able to observe assemblies that correlate with the subjective experiences of, for example, neuropathic pain, depression and schizophrenia. Nevertheless, researchers have already observed the assembly model in action. In 2006 Toby Collins and others in my group at Oxford showed that in rats, the formation, activity and duration of assemblies correlate selectively with the action of anesthetics. Pilot observations in our laboratory, yet unpublished, also show that the number of neurons active in assemblies in the sensory cortex of an anesthetized rat reflects degrees of anesthesia. Earlier this year another member of my team, Subhojit Chakraborty, demonstrated that in rats, assemblies in the visual and auditory systems might serve as a good basis for distinguishing the subjectivity of seeing versus hearing.

Other criticisms relate to time and space. In epilepsy, for example, a prolonged neuronal assembly sustains a seizure, which equates with a *loss* of consciousness. But the whole point of assemblies as the appropriate NCC is that they are highly transient; a seizure acts as a jamming mechanism that prevents that transience, thus allowing a single assembly to last orders of magnitude longer than normal. Collins, Michael Hill, Eleanor Dommett and I have similarly suggested in a recent paper that anesthetics also may act as a jamming mechanism.

Another area of objection is that the assembly model does not have any spatial properties; there is no identified anatomical locus. But all too often we place far too much significance on localization as an end in itself. There is no need for a "center" for any given brain function, much less for consciousness.

A more plausible scenario would be that many different brain regions, in generating highly transient assemblies, converge as inputs to a space-time manifold. The present difficulty is that we cannot describe such a manifold using current experimental techniques. Perhaps the manifold could eventually be modeled mathematically. Such models and their interactions may be the way forward.

A final problem, and one that applies to NCC at the basic level, is how they might be harnessed to tackle the hard problem: determining how physiological events in the brain translate into what you experience as consciousness. We will not be in a position to find a solution until we know what kind of evidence would satisfy us: A brain scan, a performing rat, a robot, a formula? Or perhaps an induced change in one's subjective state, such as if Christof's brain could be manipulated so that he would experience the world as I do—and even agree with me. ■

"There is no need for a 'center' for any given brain function, much less for consciousness."

MORE TO EXPLORE

The Private Life of the Brain. Susan Greenfield. John Wiley & Sons, 2000.

A Framework for Consciousness. Francis Crick and Christof Koch in *Nature Neuroscience,* Vol. 6, pages 119–126; February 2003.

The Quest for Consciousness: A Neurobiological Approach. Christof Koch. Roberts & Company Publishers, 2004.

A Neuroscientific Approach to Consciousness. Susan A. Greenfield and T.F.T. Collins in *Progress in Brain Research,* Vol. 150, pages 11–23; 2005.

The authors would like to thank Joseph Dial of the Mind Science Foundation for his support. Susan Greenfield thanks Michael Hill, Nicholas Shea and Kathleen Taylor for their insights.