Left: Looking like a knight in a chain-mail hood, Koch dons an array of electroencephalography (EEG) electrodes for a brain-function test.
Right: Afterward, he appears to have been attacked by a giant squid, thanks to the array's suction cups.

Be Aware of Your Inner Zombie

By Andrew Porterfield

Zombies walk among us. In fact, we couldn't get along without them—operating below the threshold of awareness, zombie systems in our brains take care of all sorts of routine tasks without any conscious effort on our part. Studying such unconscious processes is beginning to throw light on how the conscious mind works.

In the 1968 horror classic *Night of the Living Dead*, terrified people trapped in a Pennsylvania farmhouse try to survive zombies hungry for human flesh.

But real zombies aren't like that, according to neurobiologist Christof Koch. "The word 'zombie' is a surprisingly technical term, developed in detail by philosopher David Chalmers. Zombies are exactly like you and me except that they have no feeling or awareness," Koch says—a rather more sympathetic view than director George Romero's. Chalmers's 1996 book, *The Conscious Mind*, proposed these zombies as a "thought experiment" through which we could explore the question of whether a creature could exist that displayed the full range of human behavior but lacked conscious sensations. Such a zombie would get up, get dressed, and go to work like you and me. If you asked it over lunch what its favorite band was, it might answer, "Pink Floyd," and perhaps even invite you on a date to see a local tribute band cover *Dark Side of the Moon* on Saturday. But the zombie would not be experiencing the taste of the sandwich it was eating, nor would it "enjoy" the music as a human would. Do any natural laws prohibit the existence of such beings? Chalmers asked.

In fact, it's the unconscious, or "zombie," systems in our brains that help us get through daily life, Koch says, and they can show us how consciousness really works.

Koch has been fascinated by the phenomenon of consciousness for more than two decades, but his interest started not with a penchant for horror cinema, but with a toothache. "I was teaching a course at the Marine Biological Laboratory in Woods Hole, Massachusetts. So I was lying in bed, had this terrible toothache, I'm taking aspirin but it's still persistent." He began to ask himself, "Why should that hurt? Where does the feeling come from?"

Koch blends techniques from psychology,

biology, and neurology to attack the fundamental questions of consciousness: "What is it in the brain that enables us to feel? What part of the human and animal brain is necessary to be conscious? And how does consciousness arise out of matter?"

The debate about whether there's an actual place for consciousness in the brain goes back millennia. Plato and Aristotle held that the mind was entirely divorced from the physical body—like parallel Las Vegases, what happened in the body stayed in the body; what happened in the mind stayed in the mind. More recently, the French philosopher René Descartes argued that the mind and the brain could influence each other. He proposed that the soul resided in the pineal gland—a solitary lump in the shape of a pine cone (but about the size of a grain of rice) that lies almost exactly in the middle of the brain, surrounded by the matched pairs of structures that make up the rest of the brain. This gland's singular nature and central location, he argued, clearly marked it as something special, and what could be more special than the physical seat of what makes us human? (Alas, we now know that the pineal gland doesn't do much more than help control our cycle of waking and sleeping by secreting the hormone melatonin in response to darkness.)

A RISKY BUSINESS

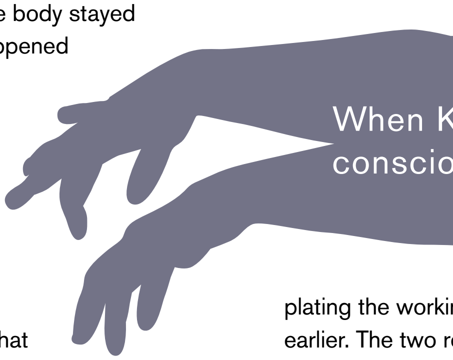
When Koch entered the field in 1988, the study of consciousness was not considered serious science. A mountain climber and trail runner, he describes himself as a risk taker. "Even without tenure, I was adventuresome, and I was very interested in consciousness. But talking about consciousness was a sign that you were retired, or had a Nobel Prize, or were a mystic and slept next to crystal pyramids," he laughs. "It was like talking about sex during Victorian times—it was just taboo." Koch paired up with the vigorously *un*retired Nobel laureate Francis Crick, the codiscoverer of the double-helical structure of DNA, the molecule of heredity, who had moved from probing the workings of genes to contemplating the workings of the brain 15 years earlier. The two remained close collaborators until Crick's death in 2004.

"Francis had thought for a long time about consciousness, and in his view it was a scandalous state of affairs that people were studying the brain without ever referring to the fact that this brain produces consciousness every day, day in and day out," Koch recalls. Together, the duo published about two dozen papers; Crick also wrote the foreword to Koch's 2004 book, *The Quest for Consciousness*, which introduced the notion of zombie agents to the unsuspecting world beyond the neurobiology community.

Today, many laboratories study consciousness, and Descartes's intellectual descendants are still asking: Where in the brain is the mind? Consider a computer, Koch says. "If you rig a thermometer to a computer and put them both next to a heat source, the thermometer will transmit the ambient temperature to the computer. Once the temperature rises above some threshold, the computer can be set to print out a message. 'It is too hot. I'm in pain.' But does the computer actually experience pain? I don't think so. At this point, its response is nothing but a programmed instruction, a reflex. At a certain threshold temperature value, some electrons flow onto a gate, a transistor opens another gate, opens the register, records the content, and prints out a statement. There is no feeling involved."

Now compare that to the sensation of someone stomping on your toe. Again, a train of electrical impulses—this time mediated by calcium, sodium, potassium, and other ions—flows from your foot to your brain. "I can say this is just biophysics—signals moving about inside the nervous system. There's no pain feeling anywhere. But without any doubt you'll feel this bad thing, pain. So how is it that a physical system like a brain can produce a subjective state, but another physical system, the computer, doesn't? This is the heart of the mind-body problem."

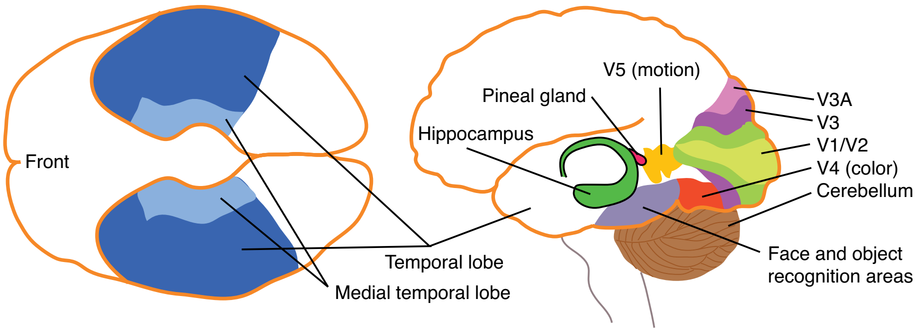
Some headway has been made since the days of Descartes, and scientists have discovered that not all of the brain is involved in creating consciousness. The cerebellum, for example, controls the timing of motor movements, and it contains half of the brain's neurons. If you lose your cerebellar function, you will be unable to coordinate your muscle movements, a condition called ataxia. You might stagger and sway while you walk, for instance. "You won't be in a rock band, you won't be a ballerina or a climber, but your visual consciousness will be marginally impacted, if at all," Koch notes.



When Koch entered the field in 1988, the study of consciousness was not considered serious science.

In Koch's world, the zombies are in our brains. In the pop-culture world, the zombies are out for our brains.

Nudist Colony of the Dead, © 1991 Pirromount Pictures.



Left: The regions mentioned in this article lie deep within your brain. (The brain's outer surface is outlined in orange.) Visual processing begins in area V1, also known as the primary visual cortex. V2 through V5 do additional analysis before sending the information on to the medial temporal lobe, the hippocampus, and other areas where consciousness may lurk. V1 through V5 are on the inner, facing surfaces of the cerebral hemispheres; the hippocampus, the cerebellum, and the pineal gland straddle the brain's midline.
Far left: A horizontal slice through the temporal lobes.

I SEE, THEREFORE I AM

Koch attacks the problem of consciousness through the brain's visual system. He originally established his laboratory—not coincidentally, painted in bright primary colors—to study visual perception, trying to understand how we focus on one aspect of a scene, and to figure out how this form of selective visual attention could be taught to machine-vision systems. This background lent itself easily to experiments with visual consciousness.

Koch's other longtime collaborator, Itzhak Fried, is a professor of neurosurgery at UCLA who implants microelectrodes into epilepsy patients whose seizures cannot be controlled by nonsurgical means. Fried uses these electrodes to find the abnormal brain-cell activity caused by epilepsy; the electrodes pinpoint the lesions' locations in order to guide his scalpel. The electrodes—as many as a dozen per patient, with each one sprouting as many as nine microwire probes from its tip—show up nicely in CAT scans, which provide their tips' three-dimensional coordinates to within a few millimeters.

Even better, from Koch's point of view, each microwire is sensitive enough to pick up the musings of a single neuron. As luck would have it, some of the deep-brain centers involved in recognizing and remembering people are in brain areas that are most often affected by epilepsy, such as the hippocampus. Thus Fried's electrodes give Koch a window into—or, more accurately, a water glass pressed up against the wall of—those regions of the brain.

The electrodes remain in the patients' heads for up to two weeks. That's a long time to spend hanging around in a hospital, so the simple video games that Koch and his colleagues have designed for their experiments offer a welcome distraction.

As the patients stare at the computer screens, the researchers use the electrodes to look for neural correlates of consciousness, or NCCs, which Koch and Crick had defined in a 2003 paper called "A Framework for Consciousness." The paper proposed that visual NCCs are small coalitions of nerve cells that collect information from the back of the cerebral cortex, where the preliminary processing of visual information is performed, and establish sets of two-way communication links with other parts of the cerebral cortex at the front of the brain.

The receiving regions include the medial temporal lobe, where Koch, his grad student Gabriel Kreiman (MS, PhD '02), and Fried had already found neurons that only fired when a person was consciously perceiving an image. That discovery had been made by fitting Fried's patients with special LCD glasses that were, in effect, separate TV screens for each eye. A picture would appear on one screen and remain there so that the test subjects could clearly see it. Then a second, different picture would be momentarily flashed into the other eye—a technique known as "flash suppression," because the new picture in the second eye would suppress the perception of the old image in the first eye. In other words, the fresh image wiped the older one from consciousness, even though both were still there to be seen. The neurons' firing rates reflected this. Kreiman noticed that, in most cases, each neuron being recorded would respond to only one specific image, say a picture of a smiling girl. When the image was flash-suppressed, however, the neuron became far less active, even though its preferred picture was still displayed. Those neurons, therefore, followed whatever was in the patient's conscious perception. They fired when the patient saw the image, and they didn't fire when the patient didn't. Wherever the brain's representations of the suppressed images—visible, but not consciously seen—might reside, they had to be somewhere else.

SEEING JENNIFER ANISTON

That some neurons fired only in response to specific pictures was no big surprise—after all, pattern recognition is one of the things our brains do best. But in 2005, postdoc Rodrigo Quian Quiroga; grad student Leila Reddy (PhD '05); Kreiman, by then at MIT; Koch; and Fried announced the discovery of individual neurons in the medial temporal lobes that recognized specific *people.* It didn't matter whether the picture presented was full-face or in profile, or even a line drawing or a caricature; the neuron "knew" who it was looking at.

"The first neuron we found behaving this way was a Bill Clinton neuron back in 2002," Koch recalls. "Then, there was a second neuron that responded to three different cartoon images of characters from *The Simpsons*, and a third selective neuron to basketball superstar Michael Jordan. When we first submitted the paper to *Nature*, the referees didn't believe this unheard-of

Your brain wires up groups of cells that respond to things we constantly encounter. "I had no idea who Jennifer Aniston was before we did these experiments, but presumably I now have a set of neurons that respond to Jennifer Aniston," Koch chuckles.

degree of selectivity, since we only had three such neurons. Three neurons don't make a discovery. So we went back and characterized many more of these remarkable cells—51 in that 2005 *Nature* paper, and more since then."

Grad student Stephen Waydo (PhD '08), on loan from control and dynamical systems professor Richard Murray (BS '85), and Koch pieced together an explanation for the process behind such extreme selectivity. Although the world around us presents an infinite variety of stimuli, Waydo and Koch assumed that most of the patterns that come to us through our senses are due to a small number of causes. "For instance, when I'm at home, most of the visual activity in my brain at any given moment is caused by me seeing my family and the furniture in the rooms around me, all of which are very familiar to me," Koch says. Working from this premise, Waydo devised a set of machine-learning rules that would enable a computer to identify such commonly occurring patterns—discovering for itself the Platonic form, if you will, of Koch's sofa—and then represent each one as a specific pattern of outputs from a collection of "neurons."

Similarly, says Koch, "Your brain wires up groups of cells, what we call concept cells, that fire specifically in response to things we constantly encounter." To illustrate this, he brings up on his computer a session with one of Fried's patients. An image of Marilyn Monroe appears on the screen: a rapid-fire *trrrppp*, *trrrppp*, *trrrppp* pours from the speakers. Then, actor Josh Brolin; nothing.

"I had no idea who Jennifer Aniston was before we did these experiments, but presumably I now have a set of neurons that respond to Jennifer Aniston," Koch chuckles. "It's an efficient way of dealing with the world. It allows infants to learn early the lessons that stay with us: first you learn to recognize your parents and your siblings in this abstract and invariant matter, and your dog, and all the other important people, animals, and things that your brain constantly encounters. Then when you get older, it's on to mastering more abstract things, like Marcel Proust or $e = mc^2$."

Concept cells respond to sensory stimuli of all kinds—in Aniston's case, for example, not just seeing her, but hearing the sound of her voice or even reading her name; this set of neurons will activate when exposed to any aspect of the Zen of Jen.

There are two schools of thought about how concept cells work. The distributed-population hypothesis invokes a large number of neurons, each contributing a little bit to encoding the percept. The power of this approach lies in the great number of distinct objects that can be encoded, and in the robustness of their representation—lose any one neuron, and the percept hardly changes. The sparse-coding hypothesis, on the other hand, proposes that a small network of neurons is entirely responsible for the encoding. The ultimate sparse network would be one consisting of a single cell; this reduction to the extreme is known in the trade as the "grandmother cell" hypothesis, because it implies that somewhere in your brain there lives a cell whose sole duty is to recognize your grandmother.

At first blush, the existence of Jennifer Aniston neurons would seem to support the grandmother-cell hypothesis, but "there is something to both sides," says Koch. In 2008, Quian Quiroga, Kreiman, Koch, and Fried found that yet another of Fried's patients had a neuron in the hippocampus (an interior region of the medial temporal lobe) that responded not only to Aniston but to her *Friends* costar Lisa Kudrow. Since the two actresses have only one degree of separation on the small screen, it seems reasonable that their NCCs might share a few cells as well. (Another cell in the parahippocampal cortex of a different patient fired in response to both the Eiffel Tower and the Leaning Tower of Pisa but not to other landmarks, displaying a similar power of generalization.) So it appears that the networks are indeed sparse, but maybe not *that* sparse—they contain enough members to be both very selective and very abstract at the same time.

A couple of famous people who might be in your head.
The first "concept cell" that Koch's team discovered fired in response to pictures of President Clinton.
If you've watched more than a few episodes of *Friends*, you've probably got a set of neurons that respond to actress Jennifer Aniston.
Once a network of concept cells has been wired up, it will be activated whenever it recognizes the object of its obsession. It doesn't matter whether the image it sees is a grainy black-and-white photo or even a scrawled cariacture; nor does it matter if the subject is seen in a full-face view, in profile, or even partly obscured.
In fact, the stimulus doesn't even have to be visual. Concept cells will react to sounds, smells, and even the written word—any sensory stimulus that we associate with that person.

LIVING THE ZOMBIE LIFE

So what do zombies have to do with any of this? More than we'd like to admit, apparently. In his upcoming book, *Consciousness—Confessions of a Romantic Reductionist*, Koch recalls a trail-running session in which he encountered a rattlesnake. "Something made me look down. My right leg instantly lengthened its stride, for my brain had detected a rattlesnake sunning itself on the stony path where I was about to put my foot," he writes. "Before I had seen the reptile or experienced any fear . . . I had already acted to avoid stepping on it." Had he been forced to think about consciously adjusting his stride, it would have been too late.

This unconscious, automatic response was choreographed by one of our zombie systems. While it's well known that the nervous system controls many body functions without conscious effort—things such as heartbeat, breathing, and digestion—Koch estimates that about 90 percent of our activities are the work of unconscious zombies.

"The central insight of Sigmund Freud is that you're not conscious of most of the stuff in your brain. For example, we spend much of the day typing. Now, if I ask someone, 'What finger do you type the letter *f* with?' most people won't know. They have to pantomime the movement to realize that it's the left index finger. But if you don't consciously think about it, your fingers will do the typing by themselves." Even activities as seemingly varied and unstructured as Koch's mountain-climbing adventures call on these zombie systems most of the time. Nearly all of his risky moves up the side of a cliff are so ingrained that he doesn't give them a thought.

"So why isn't all of life like that?" he asks. "Why not have a completely zombie existence?" Because life throws us curve balls, that's why. "The world is so complex; you have to do things that are nonroutine. Let's say there was an earthquake right now. You would look first at the glass window, which could shatter and seriously injure you, and then you would look around for a safe way to get outside. Reacting to an earthquake isn't something you've trained for hundreds of times." But for repetitive behaviors, even very elaborate ones, it's a convenient way for our brain to handle the situation with minimal effort.

Could the complexity of a system automatically create consciousness? Are we on the verge of a sentient Internet? Maybe, Koch says.

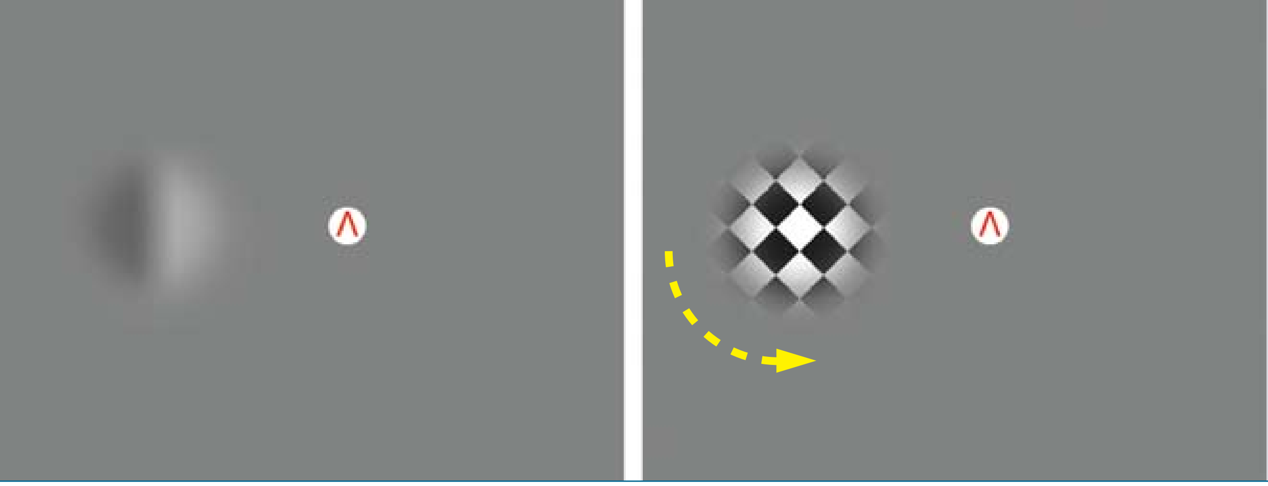
PAYING ATTENTION

But even with an army of zombies at our command, "we suffer from information overload," says Koch. "We have to concentrate on the essentials; otherwise we wouldn't get anything done. Attention is the brain's way of maintaining focus." But attention is not the same as awareness or consciousness, he adds. While consciousness involves the general awareness of the world around us, or what we think is the world around us, attention is a spotlight. Attention takes hold of one aspect of our environment, whether it's scanning your DVD collection for the seventh season of *Friends* or listening to one person in a crowded room. It's a mechanism for selecting for further processing a few rivulets of information out of the flood that inundates our senses, providing the brain a way to organize multiple inputs and make sense of the world.

Scientists have long assumed that attention and consciousness are the same, or at the very least heavily intertwined. This past May, biology postdoc Jeroen van Boxtel, psychology and neuroscience postdoc Naotsugu Tsuchiya, and Koch demonstrated that this assumption is wrong. In these experiments, members of the campus community were asked to fixate their eyes, without any movement, on a dot in the center of a computer screen. Then, off to one side of one eye's field of view, a Gabor patch—a computer-generated blur resembling a small smudge—would appear for four seconds and then vanish, leaving an afterimage in the eye that had seen it. (Afterimages are the oldest tools of visual psychologists, as they are easy to induce and manipulate in reproducible ways.) The volunteers pressed a button when the afterimage disappeared.

At the same time, the participants were asked to count the number of times a specific symbol appeared among a series of symbols that rapidly flashed, one by one, at the center of their gaze. In some of the runs, the correct symbols were made deliberately hard to spot—a demanding task requiring full attention. In other runs, the task was easier, meaning that the volunteers could divide their attention between the Gabor patch and the stream of symbols, even as they continued to stare at the dot. Either way, the subjects were conscious of the patch, regardless of how much of their attention they could give it.

But here things got interesting. In a second set of experiments, the Gabor patch was removed from conscious perception by using the flash-masking technique: a



The two-eye test of consciousness. As the subject stared at the white disks, a Gabor patch would appear in one eye's field of view (far left). If a rotating checkerboard (left) was then flashed in the other eye to suppress the Gabor patch, the patch's afterimage was also affected.

second, high-contrast pattern that flickered and rotated was shown in the same relative location as the patch, but in the other eye's field of view. Now the subject could see only the moving pattern, not the stationary Gabor patch.

Not surprisingly, the afterimages lingered longer when the Gabor patch was consciously visible. One would also expect that the afterimages would be more persistent when attention was being paid to the patch, because of the mental effort devoted to processing that visual information. In fact, the opposite occurred. When the subjects had to pay full attention to the symbols at the center of their gaze and therefore couldn't concentrate on the patch out at the periphery, the patch's afterimage took *longer* to disappear—and this was true no matter whether the patch was masked or visible.

For the first time, consciousness and attention had been teased apart and shown to operate not only independently but in opposition to each other—a percept had been affected by whether or not it had been a focus of attention, regardless of whether the subject had been conscious of seeing it. This implies that somewhere in the brain, focused attention and consciousness—without attention—are somehow being handled differently. "The history of any scientific concept—energy, the atom, the gene, cancer, memory—is one of increased differentiation and sophistication until it can be explained in a quantitative and mechanistic manner at a lower, more elemental level," says Koch. "Making the distinction between attention and consciousness clears the decks for a concerted, neurobiological attack on the core problem of identifying what is necessary for consciousness to occur in the brain."

DESCARTES AND SOUL

Religion is rarely a popular topic in the laboratory, but anybody who studies consciousness can't escape the question: where, if anywhere, is the soul? Are we something more than a mosaic of cells, proteins, lipids, and DNA?

"It's a very old concept, and it means many things to many people. This is what Descartes, among others, was addressing," Koch says. "Consciousness is definitely the modern conception of the soul. But does that mean the soul resides in the brain?"

So, while Koch and his team are focused on finding the neural correlates of consciousness, and mapping the exact pathways that give rise to awareness, he's looking to what lies beyond. "It may be the complexity that matters. Consciousness is a property of complex entities and cannot be further reduced to the action of more elementary properties," he writes in *Consciousness*. Could the complexity of a system automatically create consciousness? Once a technological threshold has been crossed, could we re-create it? Are we on the verge of creating a sentient Internet? Or a robot that can feel?

Maybe, Koch says. If we can define consciousness well enough to pick out an NCC and say, "That's how we feel pain," we might be able to create a machine that "experiences" the same sensation. That done, we could download the electronic NCC onto a disk, and, as Koch cheekily proposed in a 2008 *IEEE Spectrum* article, auction it off on eBay. Talk about selling your soul. E&S

Christof Koch is the Troendle Professor of Cognitive and Behavioral Biology and professor of computation and neural systems (CNS). He studied physics and philosophy at the University of Tübingen in Baden-Württemberg, Germany, earning his MS in 1980 and a PhD in biophysics in 1982. He came to Caltech as an assistant professor in 1986 to join the just-established CNS program—the first of its kind in the world.

On November 18, Koch and Caltech biologist David Anderson were selected as two of the inaugural Allen Distinguished Investigators by the Paul G. Allen Family Foundation—a group of seven scientists "working on some of the most exciting research in biology and neurology," according to Microsoft cofounder Paul Allen. ***(For more on Anderson's work, see page 25.)***

Koch's research has been funded by, among others, the National Institute of Mental Health, the National Science Foundation, the Defense Advanced Research Projects Agency, the Office of Naval Research, the Gordon and Betty Moore Foundation, the Swartz Foundation for Computational Neuroscience, the G. Harold & Leila Y. Mathers Charitable Foundation, and the Gimbel Fund.

This article was edited by Douglas L. Smith.