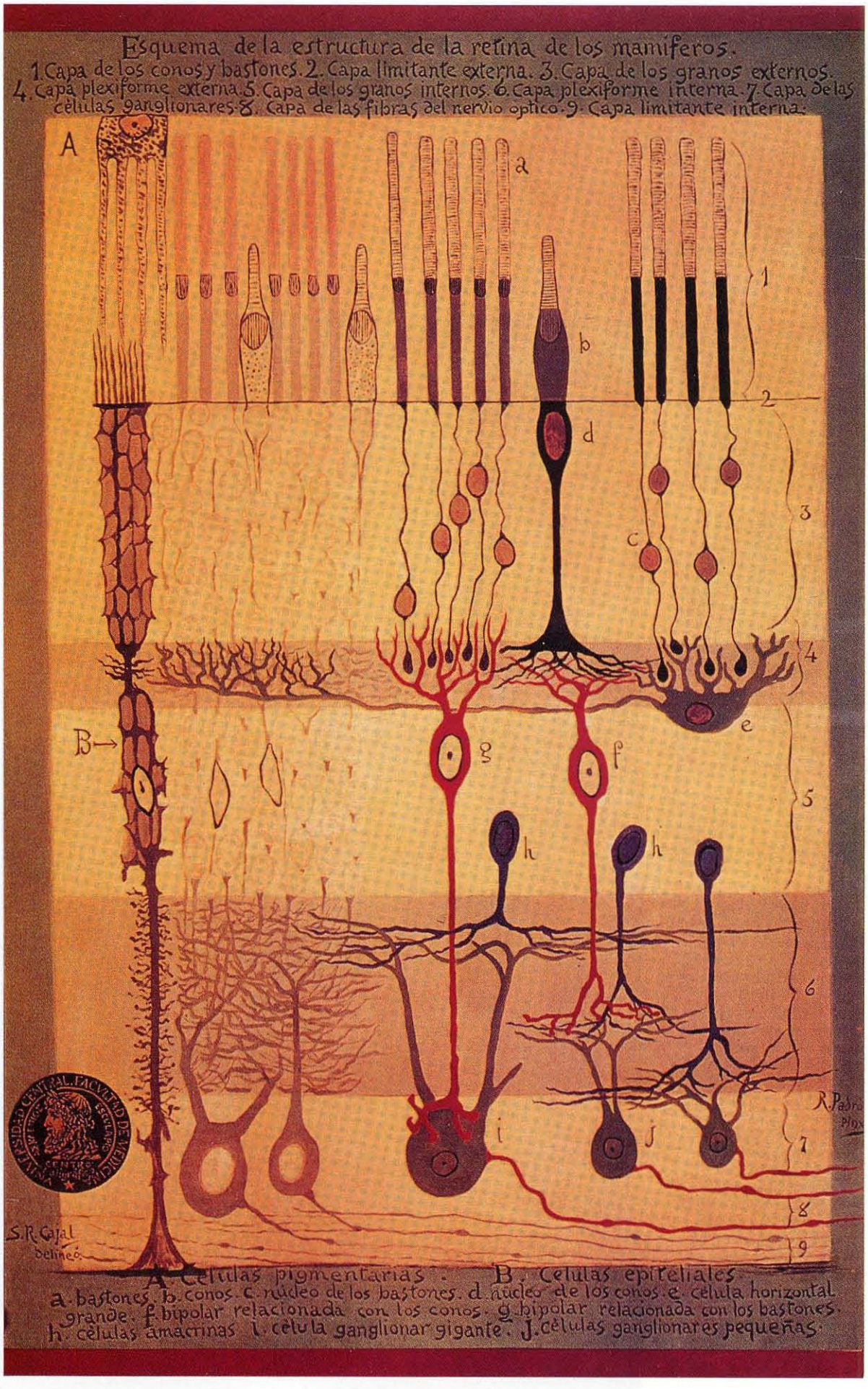

Esquema de la estructura de la retina de los mamiferos.
1. Capa de los conos y bastones. 2. Capa limitante externa. 3. Capa de los granos externos. 4. Capa plexiforme externa. 5. Capa de los granos internos. 6. Capa plexiforme interna. 7. Capa de las células ganglionares. 8. Capa de las fibras del nervio optico. 9. Capa limitante interna.
A
B
a
b
c
d
e
f
g
h
h
i
J
1
2
3
4
5
6
7
8
9
S.R. Cajal delineó.
A. Células pigmentarias. B. Células epiteliales
a. bastones. b. conos. c. núcleo de los bastones. d. núcleo de los conos. e. célula horizontal grande. f. bipolar relacionada con los conos. g. bipolar relacionada con los bastones. h. células amacrinas. i. célula ganglionar gigante. J. células ganglionares pequeñas.

When Looking Is Not Seeing: Towards a Neurobiological View of Awareness

by Christof Koch

I'll try to explain how we can begin to attack the problem of consciousness in a reductionist, scientific manner.

Three fundamental problems intrigue scientists today. The first is the physicist's dream, namely, to unify all the known forces in the universe into one single theory, and there's a candidate being worked on here at Caltech called superstring theory. This theory—whatever it proves to be—together with the initial conditions prevailing at the birth of the universe 15 billion years ago, would explain why the universe is in the bad shape it's in now. The second problem is the biologist's dream—to explain how a single cell, over five weeks, five months, or five years, becomes a plant, a bug, or a person—the problem of development. There are lots of people at Caltech working on that problem, too. The third has been the domain of philosophers and psychologists until now. It's the problem of the brain—how do we perceive? Not only we humans, but monkeys, cats, and even such lowly creatures as the fly and the sea slug—animals we squish underfoot sometimes. How do we perceive, and react to, our surroundings in a way that makes sense?

Where seeing begins. This cross-sectional drawing of nerve cells in the retina was made by Santiago Ramón y Cajal, who shared the Nobel Prize for medicine with Camillo Golgi in 1906 for their studies of the nervous system. From the light-detecting rod and cone cells at the top to the optic-nerve fibers at the bottom, the retina is about 0.01 inch (0.25 mm) thick.

Solving this problem is really preliminary to solving the problem that drew many of us into neurobiology in the first place, but which we can't talk about. It's the evil C-word, where C stands for consciousness. Over the last 60 years, particularly in this country, there has been a very strong movement by the behaviorists—B. F. Skinner and friends—to outlaw consciousness. They say that consciousness is not really a scientific concept. You can't test it, so you should just leave it out of your experiments altogether. But we all know that we *are* conscious, so I'll try to explain how we can begin to attack the problem of consciousness in a reductionist, scientific manner.

There are some house rules to this game, in order to not get stuck early on. The first one is: don't attempt any formal definition of "consciousness." We roughly know what we're talking about, and for any definition you give describing a "conscious" being, I can give a counterexample involving sleepwalking, or REM sleep, or anesthesia, or zombies, or something. So, without defining it more precisely, "consciousness" is the state that I hope you're in now. You aren't asleep yet—that may come as you read—and that's the state I mean.

Rule two is that we are going to assume that higher animals—particularly monkeys, but probably also cats and dogs—have some form of consciousness. If you look at the brain structure of our closest cousins, the great apes, it is very similar to ours. Our brain is bigger, but the complexity is comparable. There's no reason to assume that they don't share some degree of the consciousness that we have. It's a corollary of this rule that a language system is not required for consciousness. From there, it becomes a question of which animals are not conscious, and that again is best left for when we know much more about consciousness. The sea slug, I would say, is probably not conscious, but where consciousness begins is diffuse.

And I'll disappoint you with rule three. I'll not deal with the most interesting aspects of consciousness—such things as free will and qualia. Qualia are subjective properties such as

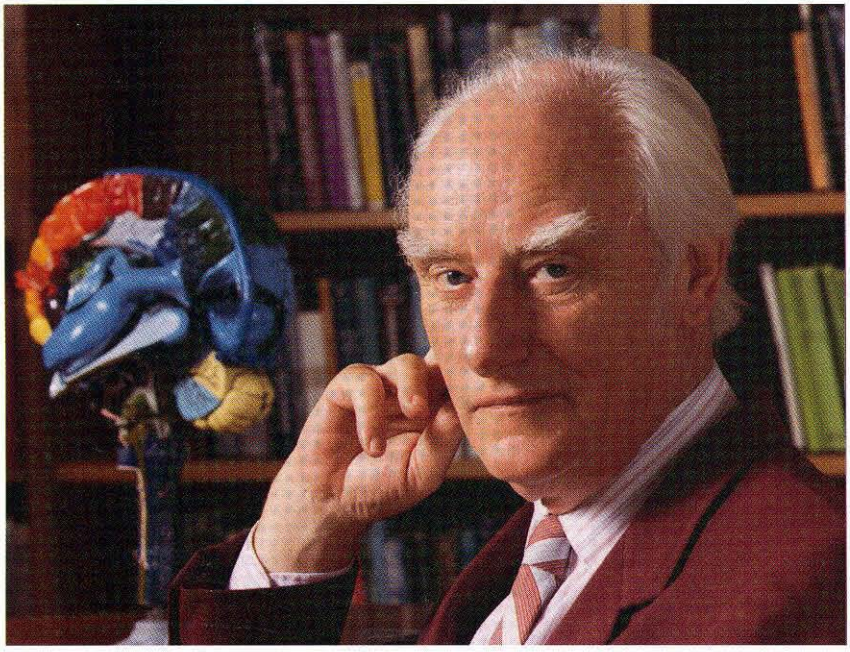
blueness or pain, considered separately from the objects that manifest them. If I have a really bad toothache—I've already taken two aspirin and I'm lying in bed and the tooth starts pounding away—I might start thinking, "Why is this actually *bad?* Why is this awful feeling associated with pain?" The pain comes from a nerve cell—a neuron—in my tooth sending a message to my brain, producing electrical activity in some brain cell. But why should that activity be bad? If a neuron two inches away is activated, it causes me pleasure. And if a neuron two inches in the other direction is activated, I smell a rose. How can the electrical activity of any cell, how can any physical cause, give rise to these feelings of blueness, of awfulness, of pleasure, of smelling a rose? This is a very subjective question, of course, because the feeling *I* have of pain might be quite distinct from the feeling *you* have of pain. All I know is if I hit you on the foot with a big hammer, you are going to cry out. I can infer from your reaction that you probably have the same awful feeling that I would have under the same circumstance, but I really can't test it. And so these problems are best left out. They can only be addressed within philosophy, and may never have any scientifically testable explanation.

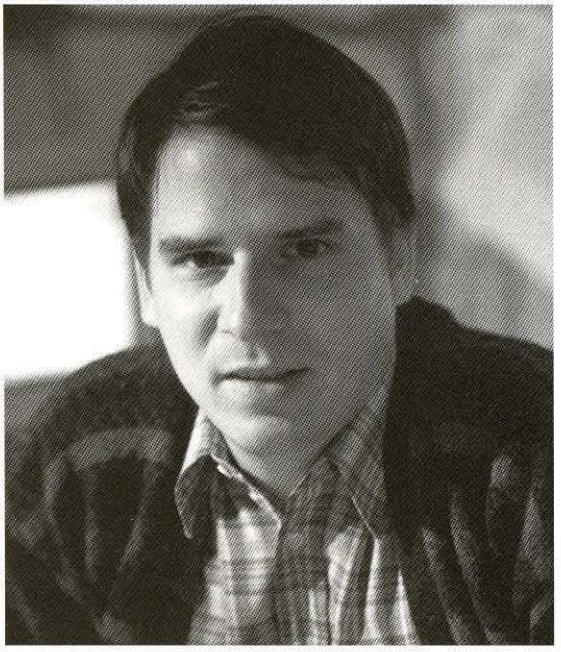
If I'm lost in some problem, I can drive home and realize suddenly that I'm in my garage, but I don't have any awareness of how I arrived there. Yet I had to stop at red lights, make left turns only when there was no oncoming traffic, and, in general, act in an intelligent manner.

And no amount of psychoanalysis, of lying on the couch and paying 100 bucks an hour, will ever tell me how I see color. Maybe I can uncover why I married my wife, but low-level things—how I see color, how I hear or smell—are not amenable to any amount of introspection. In fact, we probably don't have any conscious access to most of our brain. Psychological theories about consciousness or other mental phenomena sometimes have very good elements, but they work from the outside. The only way to find out how the human brain really works is to open up the black box—in this case the brain—and do experiments. You put in electrodes; you do biochemistry; you apply the entire gamut of scientific procedures.

Having said that, I will now discuss the unconscious, a notion first proposed by Nietzsche, and popularized by Freud, Jung, Adler, and others. Over the last 20 years, cognitive psychology has made great progress in understanding a variety of aspects of the unconscious mind, especially the two called "automatic processes" and "knowledge without awareness." All of us do both of these things all the time. Driving a car is a good example of an automatic process. The first time you did it, it took all your concentration. You had to consciously pay attention to everything—staying in a lane, looking in the mirror, and shifting gears if you had a manual transmission. But now, a few years down the road, you drive completely automatically—you can even be thinking of something else. If I'm lost in some problem, I can drive home and realize suddenly that I'm in my garage, but I don't have any awareness of how I arrived there. Yet I had to stop at red lights, make left turns only when there was no oncoming traffic, and, in general, act in an intelligent manner. It happens to me all the time. Another example of an automatic process is mirror writing, like Leonardo da Vinci did in his notebooks. Most people can be trained to read and write mirror writing. It's difficult, and takes quite a while, particularly writing it. But if you *do* it, if I pay you as an experimental subject, you can acquire it in a couple of months. And then you do it effortlessly—it's just like reading normal writing, which, incidentally, is also an automatic process. The other aspect, knowledge without awareness, is knowledge that's available to the brain, but not the mind—you know something, but you're not aware that you know it. One example is subliminal advertising, which was very controversial in the 1960s. The effect is not nearly as strong as most people believe, but it exists. I can flash the words *Buy My Book* on a screen so fast that you're unable to recognize them, yet something in your brain will know. But it won't make you go out and buy my book, unfortunately. A lot of the social judgments that govern our day-to-day interactions—why you like or dislike someone, why you look up to some people and down on others—have been extensively studied. They, too, bypass awareness—you like someone "instinctively," and you can't explain why.

How do we test this morass of feelings to which we have no access? Knowledge without awareness has been studied most rigorously in a class of patients who have prosopagnosia—they're unable to recognize faces. They've had a stroke, or a virus, and some part of their brain is gone. (The study of brain-injury patients has been very fruitful for neuroscience, because you can see which part of the brain has been damaged, and you can find out what mental ability has been affected, and then you can infer—if you are careful—what the missing part of the brain does.) The title character of Oliver Sacks's *The Man Who Mistook His Wife for a Hat* was a prosopagnostic. If you show him a picture of his wife of 25 years, he says he doesn't know who she is. But if you measure the skin conductance of the palm of his hand while you ask him the same question—essentially the principle on which lie detectors are based—you'll see a big change. Something in his brain has recognized her, even though he's not





Crick (left) and Koch (right).

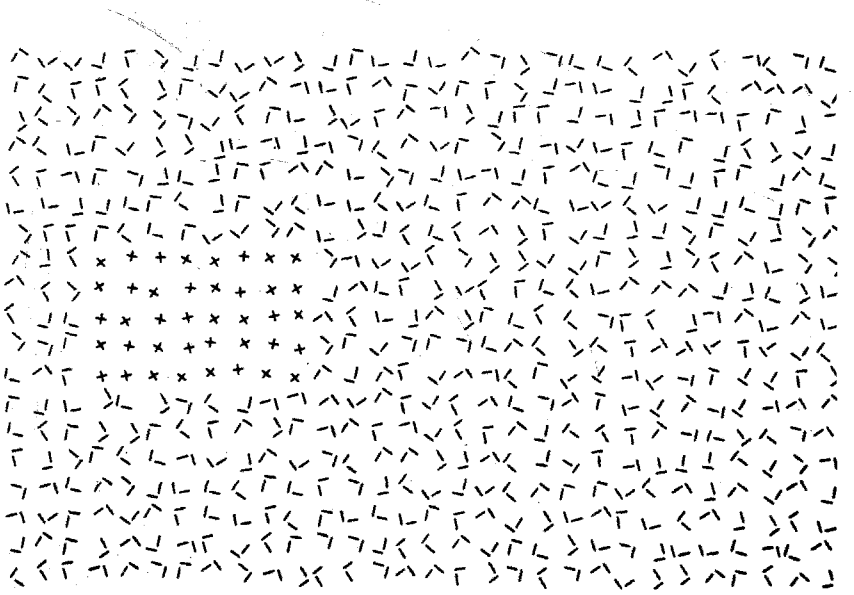
I can flash the words Buy My Book *on a screen so fast that you're unable to recognize them, yet something in your brain will know. But it won't make you go out and buy my book, unfortunately.*

aware of who she is. If you show him a random face he's never seen, he also says, "I haven't seen this before," and this time the skin conductance doesn't show any significant change. You can show him pictures of famous presidents, or movie actors, and very often you will see that, although he claims he doesn't know any of them, there's a big change in skin conductance.

There's another famous case of knowledge without awareness, from which my title, *When Looking Is Not Seeing*, comes. There's a group of patients, first discovered in England, who have what's called blindsight. They are typically older males who have had a stroke in the back of the cerebral cortex, where the visual part of the brain is, so they're unable to see anything in, say, the left part of the visual field. (They don't see anything to the left of what their eyes are focused on.) And so the doctor holds up a finger in the patient's left field of view, and asks, "Do you see anything?" And the patient says, "No, I don't. I'm blind there." "Well, do you see my finger?" "No." "Can you see it moving?" "No! Why do you ask these questions? I'm *blind!*" "Just tell me, does my finger move to the left or to the right?" Eventually the patient says, "OK, I'm just going to guess. It's moving to the left." And he's correct every time. Although all these patients adamantly insist that they don't see anything, they can correctly "guess" the direction of motion. You can move a bright light around, and they will automatically track it with their eyes, although they claim they have no knowledge of it. You can ask them to point at things, and they'll point approximately at the object, although they don't have the same visual acuity that we have. They can identify colors. They cannot do everything—for example, they can't tell shapes. If you hold up a square and ask them if it's a square or a circle, they truly seem to be guessing. This class of patient very vividly demonstrates that people can "know" something without being *aware* of this knowledge.

So what does it mean to be aware? Why am I aware of certain things and not others? How can my brain have information that I'm not aware of? Over the last three or four years, Francis Crick of the Salk Institute in La Jolla and I have outlined a framework that we think will ultimately explain this problem reductively, at the neuronal level. You probably all know of Crick, who won the Nobel Prize for his work with Jim Watson on the double helix. As for my own background, I'm a theoretical neuroscientist, not an experimental biologist. My first, and only, contact with experimental animals was when I was a programmer at the Max Planck Institute for Biological Cybernetics in Tübingen, Germany, where I later got my PhD. It was perhaps three o'clock in the morning and I was hacking some code, when a fly buzzed by with a little numbered sign glued to its back like a shark's dorsal fin. The fly had escaped from a lab where they worked on its visual system. The experience shocked me, and I've remained with computers ever since. To explain our theory, I'll proceed on three descending levels—psychologically first, on the level of the whole brain; then down to brain areas; and finally to single nerve cells.

Crick and I postulate that awareness, at the



You have to scan the image, object by object, using your mental searchlight.

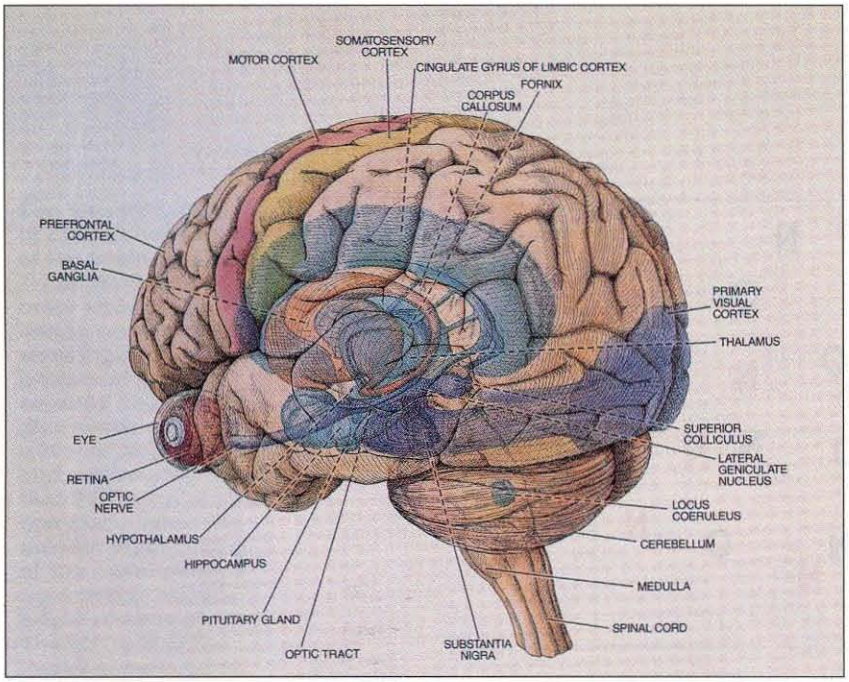
This texton pattern contains a background of L-shaped textons, in which two regions of dissimilar texture are imbedded. The region composed of +-shaped textons leaps out at you—can you find the other one?

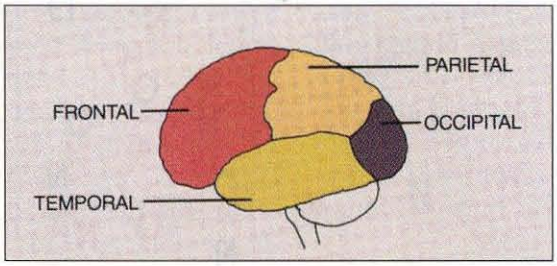
psychological level, involves two things: attention and short-term memory. These processes have been linked to consciousness for quite a while—William James described the phenomenon of attention and its relation to awareness 100 years ago. We believe that whenever you are consciously aware of something, this really means that your unconscious brain has focused your attention on that thing and put it into your short-term memory, where your conscious brain has access to it.

Attention operates in all sensory modalities, but we know it best, by far, in vision. You can think of it as a searchlight. If everything is dark, you can see only what the searchlight is shining on. Only when it illuminates the something is further analysis possible—who do you see? what are they doing? and so on. We believe something similar operates in the visual system, independent of eye movement. You can fix your eyes on one location, yet focus your attention on another; and you can shift the searchlight around. I'll try to demonstrate that with a test devised a decade ago by psychologist Ann Treisman at Berkeley. On the upper right corner of page 9, there's a figure containing the letters N and O. When you finish reading this paragraph, focus your eyes on the small drawing of the brain in the upper right-hand corner of page 7, close your eyes, turn the page, and open your eyes for just an instant—the blink of an eye, just as fast as you can make it. Then answer the question: Was there a green object? There will be many red objects, and there will either be no green objects or one green object. (The reason to glance so quickly is to avoid eye movement. It takes about a fifth of a second to initiate eye movement, so if your glance is quicker, you don't have a chance to move your eyes.) Go ahead and try it right now.

There were roughly two dozen red objects, called distracters, and there was one green target object. It turns out that the time it takes you to find this target is independent of the number of distracting objects. Whether there are 100 red objects and one green, or two red and one green, you'll still pick the green one out instantly, anywhere in your visual field. That is an example of what we call parallel processing. You don't use the searchlight of attention to do it. Even if I hadn't told you to look for a green object, you would still have seen it instantly.

Now, I'll show you something for which you need the searchlight. Using the same procedure as before, I want you to focus on the figure above on this page, turn to page 8, and look for red Os. Try it now. The length of time it takes to do this more difficult task depends on the number of distracters, so our assumption is that you have to scan the image, object by object, using your mental searchlight. So doubling the number of distracters roughly doubles the time it takes you to find the target. Bela Julesz, a visiting professor of biology at Caltech, has found that you need 30 to 50 milliseconds—roughly one-twentieth of a second—to inspect each item with your searchlight (compared to the fifth of a second it takes to move your eyes) and tell whether it's a red O. If it's not, you move your searchlight to the next target. Psychologists believe that you need the attention searchlight for this task because you need to look





Above: A mind's-eye view. The human brain consists of left and right hemispheres (the left one is shown here) divided by a deep fissure and connected by the corpus callosum. At the brain's base lies the cerebellum, which coordinates movement, and the medulla, which controls "autonomic" fucntions such as digestion, breathing, and heartbeat. The limbic system, deep within the brain, is the seat of emotion and long-term memory. Most activity related to thought and perception occurs in the brain's convoluted surface, or cortex. The primary visual cortex, where most visual processing occurs, is at the back of the brain. Right: The cortex's crumpled surface is divided into four lobes by particularly deep folds. Again, the left hemisphere is shown, in the same orientation as the large drawing.

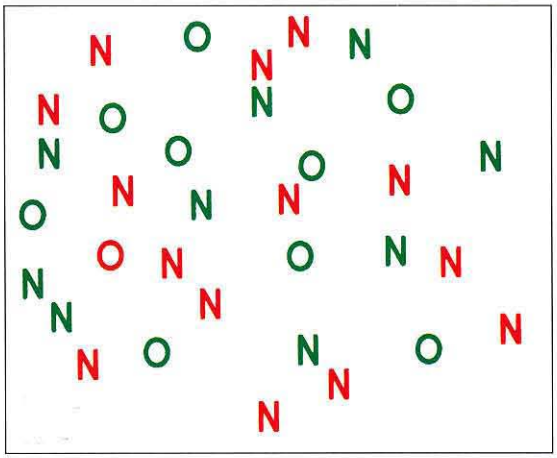
for the simultaneous occurrence of two features—the color red and the letter O.

This searchlight has nothing to do with the scanning movements your eyes make when you look at something. In the 1960s, a Russian, A. L. Yarbus, showed how people's eyes scan an object. He put a little suction cup with a mirror mounted on it on a volunteer's eyeball. The mirror reflected a beam of light onto a photographic plate, making a record of how the eye moved. He discovered that when you look at something, for instance a photograph of a face, your eyes are in constant motion. You might glance at the person's right eye first, then the left, then your gaze might move to the right ear, sweep around the edge of the face and back to the right eye again, then on to the nose, and so on. Under normal circumstances, you usually move your eyes to the same location that you move your attention, but you don't have to.

On page 6 is a demonstration that doesn't require speed. It's a texton pattern devised by Julesz. A texton is a unit of texture—a photon of texture, if you will—and in this case each texton is made up of two line segments that form either an L, a T, or a +. You'll immediately see one pattern—a region made up entirely of +'s embedded in a sea of L's—without having to scan the image. Do you see a second pattern as well? Seeing the second pattern—a region made up entirely of T's, which are texturally very similar to the L's—requires focused attention. You can't see it at a glance. So in this case you can see parallel processing and the spotlight of attention demonstrated in the same figure.

We believe something similar to the spotlight operates whenever you concentrate on one sense. You can listen to a tune, or you can close your eyes and attend to where your finger touches your leg. I'll only talk about vision from now on—that's what I know best—but the spotlight metaphor holds for the other senses as well.

Crick and I postulate that the second component of awareness is short-term, or immediate, memory. Everyone is familiar with long-term memory, which is divided into different sorts. Autobiographical memory is the one most important to us—I know where I was yesterday, or a year ago. Semantic memory is remembering facts, like what the capital of England is. These forms of long-term memory are conscious. There are also unconscious forms, such as procedural memory—skills, such as playing golf, or doing mirror writing, that you learn by practice over time. You usually don't have conscious access to procedural memory, which is why learning such skills is so difficult. The short-term memory underlying awareness is something else. If I give you a telephone number, say 359-6811, you'll remember it for a couple of seconds until something else distracts you. Or, if you need to remember it until you get home, you say, "359-6811, 359-6811, 359-6811, 359-6811." You can keep on rehearsing it indefinitely, but if you don't, it disappears. Short-term memory stores high-level information. If you're a chess player, I can show you a game in progress very briefly, and you can tell me the pieces' positions. But this is true only up to a point, because in general, this memory only holds seven things, plus or minus



Since the searchlight can inspect any one item on the tray in 30-50 milliseconds, and conscious perception takes hundreds of milliseconds, we have the subjective impression that we can be aware of all the items on the tray at once. Placing a new item on the tray causes an older item to get shoved off it.

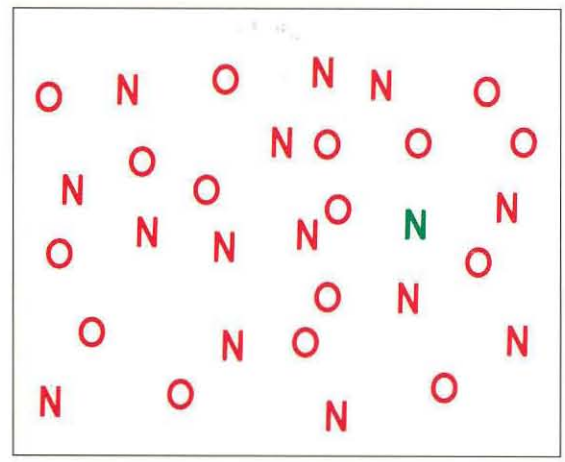
two—seven digits, seven names, seven chess positions. You could think of this memory as a serving tray with a limited capacity. The searchlight plays over the items on the tray one at a time, and as an item is illuminated, we become conscious of it. Since the searchlight can inspect any one item on the tray in 30-50 milliseconds, and conscious perception takes hundreds of milliseconds, we have the subjective impression that we can be aware of all the items on the tray at once. Placing a new item on the tray causes an older item to get shoved off it.

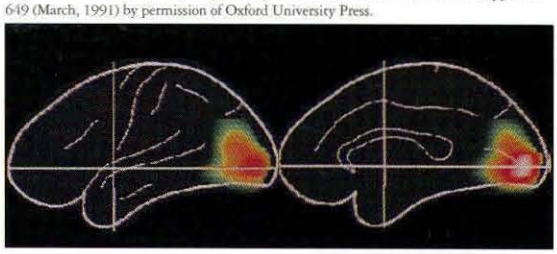
Short-term memory is very robust, and hard to damage. But there are a number of drugs used routinely during surgery that take out long-term memory. These drugs include the family of benzodiazepines—of which Valium is the best-known member—and scopolamine. But there's no drug we know of that blocks short-term memory. There are emergency-room patients who have been in serious accidents and are too badly injured to receive heavy anesthesia, so they receive benzodiazepines that relax them and induce a profound antegrade amnesia. This means that the patient now has a moving time-window of roughly two to three minutes: they forget everything that happened more than a few minutes ago, including the pain they feel during the surgery. Yet they can respond meaningfully to the requests of emergency-room personnel and can sometimes even talk, so they are conscious in the normal sense of the word, but when the drug has worn off 45 minutes later, they don't remember a thing. And there are patients who've lost their autobiographical and semantic memory systems (these two together are known as the declarative memory system) due to cancer, or surgery, or Alzheimer's disease, or an epileptic seizure. There's a patient called H. M., both of whose temporal lobes were surgically removed in the 1950s as a treatment for profound epileptic seizures. His last explicit memories are of events that happened before his operation, well over 30 years ago. He's been in the hospital ever since, and he still doesn't consciously remember his nurses and doctors. But he's perfectly aware and lucid. So long-term memory enriches our lives incredibly, but you don't need it to be aware. All that's necessary for base-level awareness is short-term memory and attention.

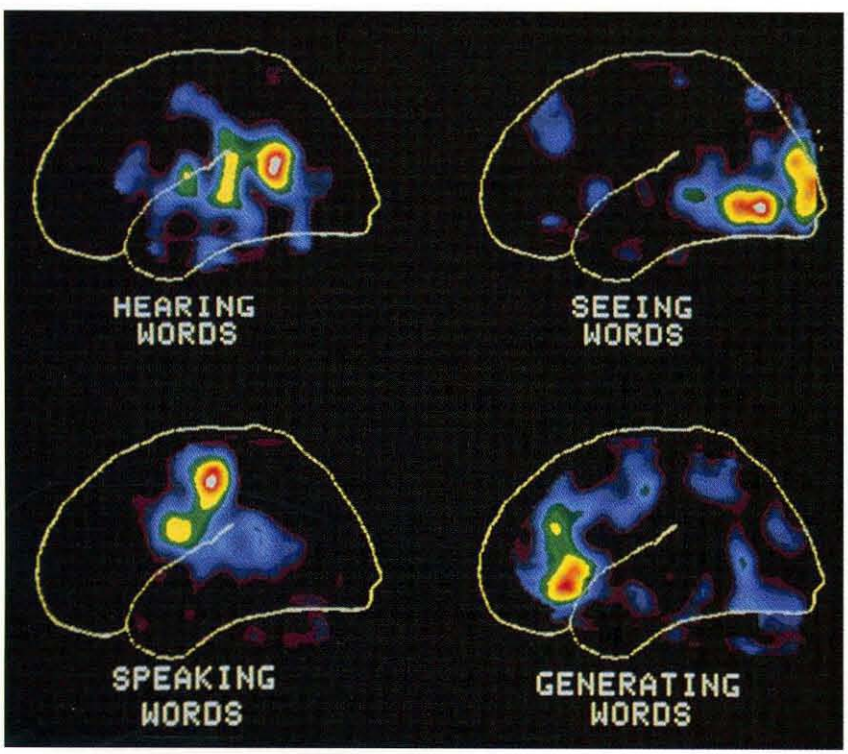
Now I'm going to show you what it means to be aware at the level of brain areas. The brain is made up of many dozens of subparts, called cortical areas, that range in area from the size of your thumbprint to a credit card. Each cortical area has a different function—for seeing color, for seeing depth, for hearing, for talking, for storing people's names, and so on. On the magazine's cover you saw a combined PET/MRI image of the brain of Caltech's John Allman, the Hixon Professor of Psychobiology and professor of biology. When you superimpose a PET image on an MRI image, you can see which brain structures are active in a particular task. In this case, John was looking at a flashing visual stimulus. The first area activated, upon arrival of visual information from the retina, is located at the back of the brain in the occipital lobe—an area called V1, for "visual area one." V1 does "early filtering"—it does the first stages of the processing needed to detect

Top left: PET image of the left hemisphere of the brain, showing areas involved in color vision. The image was made by showing a subject a pattern of colored squares and rectangles reminiscent of a Mondrian painting, and subtracting from that PET scan another one made when the subject was looking at the same pattern rendered in equally bright shades of gray. The left half of the image shows areas activated on the cortical surface, while the right half shows the interior areas.

Below: PET images showing brain areas active during various verbal skills.



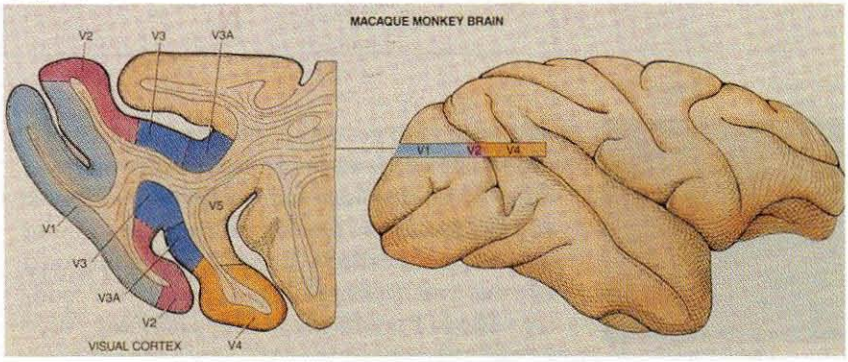


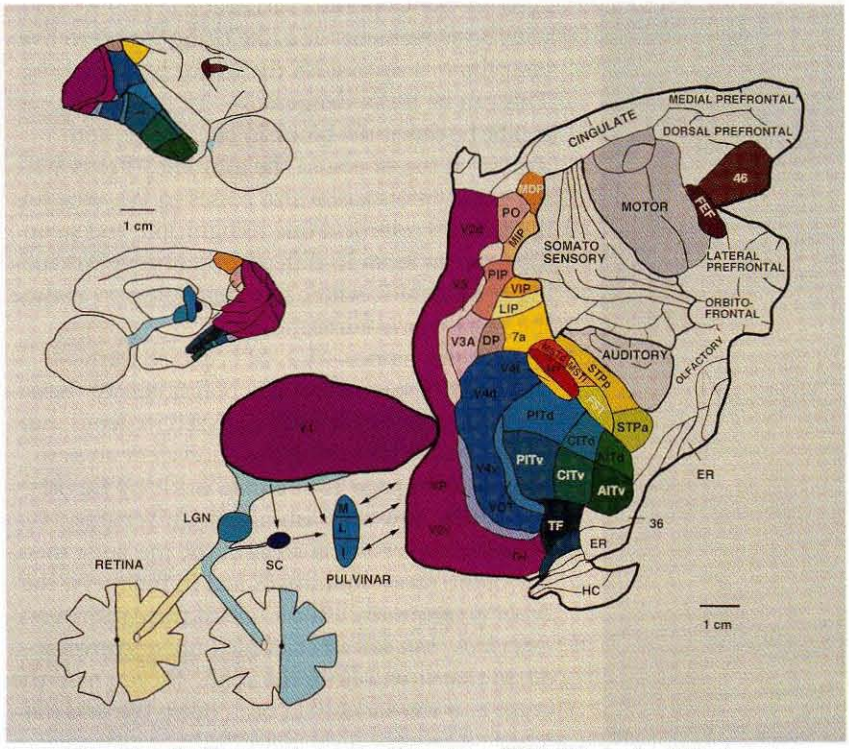


motion, and to get depth information by comparing the stereo views we get from our two eyes. From there, the visual information is distributed to other locations. One of the areas the stimulus goes to next is called V5 or MT, depending on whether you're from England or from this country. The MT area is involved in analyzing motion. Patients whose MT area has been damaged see the world as a succession of still images, with no movement. It's like being forced to live your life in a disco with the strobe light flashing. This can be really dangerous—for example, a car might be down the block in one image, and almost on top of you in the next one. From V1, the visual information also passes to V4, which is involved in color and hue recognition, and so on. There are at least 30 different brain areas, including V4 and MT, whose sole function is to analyze the visual world surrounding us.

All these names—V1, MT, and so forth—really only apply to monkey brains, where these areas' functions have been analyzed in detail, but we believe we are seeing the equivalent areas in humans. On page 10 is a map made by David Van Essen, now at Washington University, showing these cortical areas in the macaque monkey brain. Both in humans and in monkeys, the brain is essentially a sheet one to three millimeters thick, but it's all crumpled up, or convoluted, so that it will fit in the skull. So you map the brain as if the cortex had been taken out and flattened. The typical macaque brain has the surface area of one of those enormous cookies they sell in malls—160 square centimeters. Each of the two hemispheres of our brain corresponds in extent to

How a macaque sees the world. All diagrams are of the brain's right hemisphere.
Top: A cross section through the visual cortex (left) made at the level shown (right). The visual areas are shaded. Note how the cortex's deep folds mean that the brain is practically all surface. Neurons lie in the surface layers; the interior consists of connective and supportive tissues. The eyes are to the right.
Bottom: Map of the unfolded cortex, optic nerve, and both retinas. The insets show how the map relates to the hemisphere's exterior.

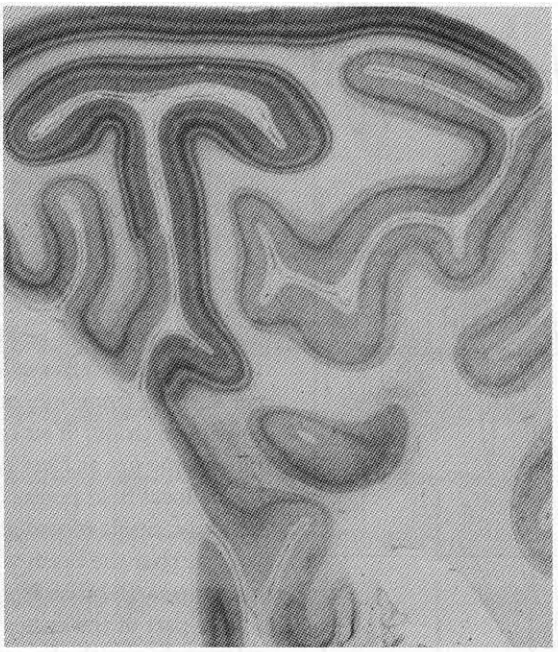


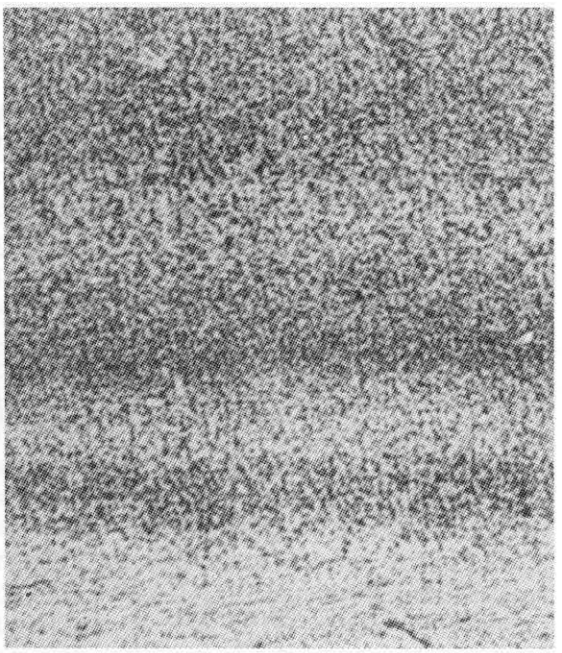


a large pizza an eighth of an inch thick and 14 inches in diameter—something like 1,000 square centimeters. And each of the cortical areas on this map contains a few tens of millions to a few billion neurons.

Let's go down one more level, and look at individual nerve cells. There are roughly a quarter of a million nerve cells and two billion synapses below one square millimeter of cerebral-cortex surface. (Synapses are the connections between neurons.) This is much more densely packed than anything we can do in silicon chips today. Seventy-five percent of those cells are pyramidal cells, so called because their cell bodies supposedly look like pyramids. Each pyramidal cell has lots of input wires, called dendrites, that branch out and extend all through the cortical layer. The cell also has one output wire, called the axon, which makes synapses with the dendrites of several thousand other cells. Neurobiologists believe that memories are encoded in the synapses, and two billion synapses per square millimeter of cortex can hold a lot of memories.

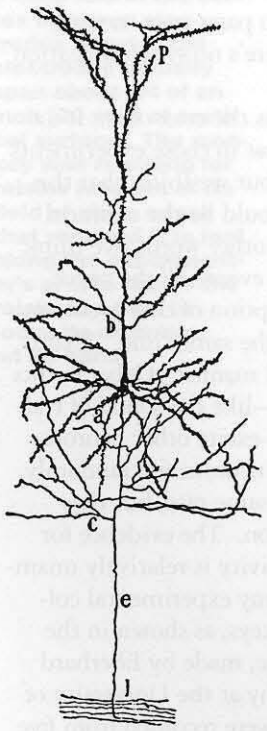
If you insert an electrode into an anesthetized animal's—or human's—brain, you can record the electrical activity of nearby nerve cells. Each nerve cell is turned on only by a particular set of stimuli—objects in the environment that the cell likes to respond to. Visual-cortex neurons like visual things. In V4, the color area, for example, there are neurons that only fire if they see objects with a reddish hue, other neurons that fire for blue, and so forth. And each neuron looks only at a small chunk of the visual field, so a specific "red" neuron will only fire when there's a red





Reprinted from D. Hubel and T. Wiesel, *Proceedings of the Royal Society*, Volume 198, pp. 1-59 (1977) by permission of The Royal Society.

Right: A cross sectional photomicrograph of the macaque visual cortex. The purple dots are nerve-cell bodies, which have been stained to make them visible. Far right: Part of the same region close up. The area photographed is about 1/8 inch (3 mm) from top to bottom.
Below: A pyramidal cell from a rabbit brain, drawn by Ramón y Cajal. The "pyramid"-like cell body lies between "a" and "b." The dendrites extend upward from "b," and the axon, labeled "e," proceeds down to the botom of the drawing, where it joins axons from other cells.



object in the part of the field it's responsible for. In a higher part of the visual cortex, certain neurons are only turned on if they see a face. So if you show a face to the monkey, assuming the face is in the part of the visual field that corresponds to where you inserted the electrode, you will see a nerve cell producing electrical activity in the form of pulses.

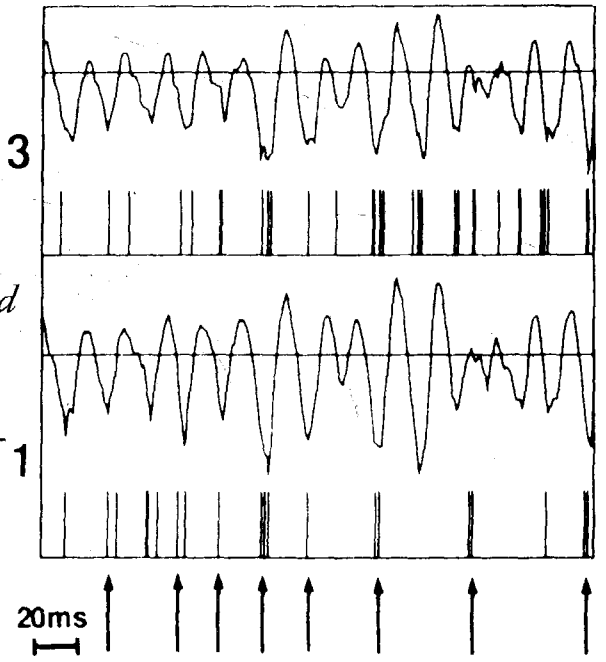
So every time I see an event, that event corresponds to electrical activity all over the brain. If I look at my friend Bill, say, his face is represented in the brain area where my face neurons are located, the hue of his face is processed in V4, the fact that he's moving around is represented in MT, my memories of him correspond to activity in the temporal lobes, and if he talks, his speech activates my auditory cortex.

Yet if I look at Bill I see a coherent whole. If he speaks to me, his voice isn't disembodied—it comes from his mouth. The color sticks with his face as it moves. And I know that it's Bill's face that's moving, not the background. How this works is a big, big mystery called the binding problem. If some perceived event outside corresponds to electrical activity all over the brain, how can I put it together in a single homogeneous image? Why don't we see the world like a cubist painting, all broken up? Nobody has ever found an area in the brain where everything comes together. A lot of people imagine that somewhere there's a control room, where a little homunculus sits in front of a TV screen so he can see, with speakers so he can hear, and he pulls the levers that make us do things. If you remember Woody Allen's film *Everything You Always Wanted to Know About Sex*, that's exactly the metaphor I mean. This control room doesn't exist—all the brain's activities are highly distributed.

The problem becomes even worse when you consider that while I'm looking at Bill, there are other people behind and next to him who are also moving and talking, and their faces and voices are being registered in these same brain areas by other neurons, yet I'm not confusing Bill or any of his attributes with those of the other people next to him. How is that possible? How come I don't get Bill's voice coming from the man behind him?

All the neurons that correspond to the object I'm attending, like Bill, must carry a common label that the brain recognizes. This label identifies for the brain all the associated neurons that are responding to different aspects of the same object out there in the perceived world. Sometimes this labeling doesn't work—for example, it's frequently a problem with witnesses in criminal cases. The witness sees something very fast—perhaps only for a tenth of a second—and remembers, "There was a man with glasses and a raincoat." And it turns out that there was one man with glasses, and another man with a raincoat, but because it happened so fast, the binding got mixed up, and a feature of one object became attached to another object. This is known as "illusory conjunction." There's also a rare clinical syndrome called disjunctive aphasia, where patients are unable to put things together. If they see two people, they'll mix up the faces, particularly if the people are the same color. Their visual fields contain two overlapping regions of

Reprinted from A. K. Engel, et al., *European Journal of Neuroscience*, Volume 2, pp. 588-606 (1990), by permission of Oxford University Press.



This synchronized oscillation could be the neuronal trace of consciousness.

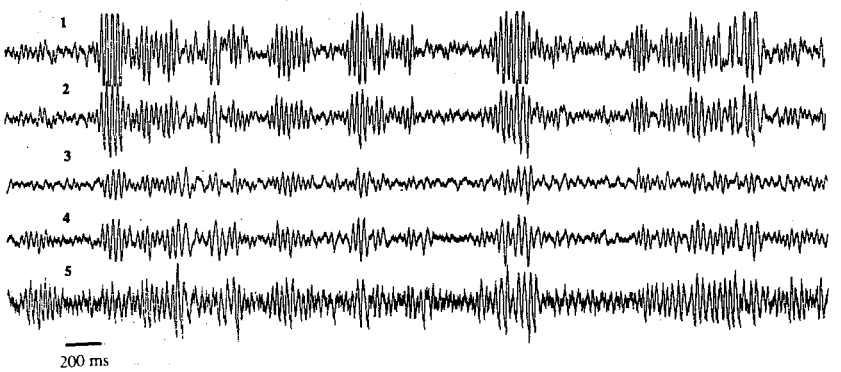
Data recorded from two electrodes, numbered 1 and 3, implanted about 0.03 inches (800 microns) apart in a cat's visual cortex. The cat was looking at a moving bar oriented 112° from the vertical. Because of the way the data is recorded, neurons generate negative peaks when they fire. The scale bar is 0.02 seconds long.

the same color, and they can't make the separation that one colored region belongs to one face, and the other region of the same color belongs to the other face. The world does look like a cubist collage to them.

What is special about the labeled nerve cells? Is there a set of special consciousness neurons—C-neurons? If so, then every time these C-neurons become activated, you're aware of the thing they correspond to. This poses a problem, because there are an infinite number of things you *could* be aware of, and this theory implies that you'd have to have a special set of neurons for each one—you'd have to have "grandmother neurons" in order to recognize your grandmother. Crick and I think that the awareness neurons are not special, but that they behave in a special way. Every neuron in the cerebral cortex has the potential to be involved, to some extent, in awareness. It's how they respond to a stimulus that matters.

Crick and I think that there's a special pattern of electrical activity that relates to awareness. It's not just that a lot of nerve cells are all firing. If you have an epileptic seizure, *every* nerve cell in your brain is firing, but you're unconscious. The above figure came from a group in Germany, headed by Wolf Singer, working together with Charles Gray, who is now at the Salk Institute. It shows the electrical activity in a cat's visual cortex when the cat saw a moving bar of light. You can think of it as being like the brain waves you see on an EEG, the electroencephalogram that doctors record. People had known about brain waves before, but not about this particular type. It's a high-frequency activity—about 40 cycles per second (cps)—and it only seems to occur under special circumstances. The two halves of the figure show the recordings from two electrodes located about a thirtieth of an inch apart—twenty to thirty cell bodies' distance, roughly. In this case, a single elongated bar of light on a dark background was moving across the receptive fields of neurons at both electrode sites. In each half of the figure, the wavy line on top shows the local field potential—the activity of several thousand nearby neurons, summed together. The series of spikes below it shows the activity from one, or a few, nerve cells closest to the electrode. You can see how the spikes or pulses line up, at least roughly, with the troughs in the local field potential. If you compare the field potentials recorded at the two electrodes, you see that the brain wave is synchronized at both sites. In other words, the electrical activity in one part of the cortex has a precise and global relationship with the activity in another part of the cortex that is responding to the same stimulus. Furthermore, the arrows at the bottom of the figure show where individual nerve cells next to the two electrodes fired at precisely the same instant. In other words, all of the neurons responding to the stimulus fire at roughly the same time. In other experiments, when the cat saw two pieces of the bar separated by a dark zone, the waves were not as well synchronized, even though the two parts of the bar were moving as one. And if the two parts start moving in different directions, there's no synchronization at all.

Crick and I think this is the crux of it. It's a bit of a leap, because the cat in those experiments was lightly anesthetized, but we think that this synchronized oscillation could be the neuronal trace of consciousness. In other words, we think that if you are aware of an event, *all* the nerve cells involved in the perception of that event anywhere in the brain fire at the same time. That is, they fire in a synchronized manner. Other events that you are not aware of—like the sound of traffic outside your window—excite other neurons simultaneously, but these neurons fire randomly. They may even fire at the same rate, but they don't fire in synchronization. The evidence for synchronized neuronal activity is relatively unambiguous in cats. Some of my experimental colleagues also see it in monkeys, as shown in the figure on the opposite page, made by Eberhard Fetz and Venkatesh Murthy at the University of Washington. The traces were recorded from five different electrodes as the monkey was taking raisins from the experimenter's hand. At each electrode, you see big waves consisting of lots of



Tracings from five electrodes implanted at roughly 1/16 inch (2 mm) intervals in the region of a rhesus monkey's brain responsible for controlling hand movements. Because of a deep fold in the cortex between electrodes 2 and 3, the electrodes actually span about 3/4 of an inch (20 mm) of cortical surface. The monkey was reaching for raisins held out of its field of view, a task that required it to feel along the experimenter's arm to locate the raisin and thus to focus its attention on its hand.

smaller spikes, and the big waves are roughly synchronized from electrode to electrode. There's less evidence in man, but a 40 cps oscillation, called an evoked potential, has been found in the auditory domain. You put two electrodes over the temporal lobes of your brain—at your temples, basically—and listen to clicks through earphones. After several hundred clicks, you'll see a few pulses of a 40-cps wave. This wave disappears in deep anesthesia. It does not disappear in sleep, and it does not disappear under light anesthesia. It is being used now in a clinical context by some anesthesiologists to check whether patients are truly under—truly anesthetized—or whether they have merely been paralyzed and rendered amnesic, somewhat like the benzodiazepine recipients I mentioned earlier.

Our theory can easily be tested experimentally. (When I say easily, I mean that conceptually it's quite simple, but actually setting it up would be rather time-consuming.) One way to do it would be to have a monkey looking at a display of red and green bars, some of which are moving to the left and the rest to the right. These figures would be very similar to the Treisman figures I showed you earlier, with the added component of motion. The fact that the bars are either red or green would cause neurons to fire in V4, and the fact that the bars are in motion would make neurons fire in MT. You train the monkey to find the odd bar—if there's only one red bar moving left, for instance—and then you look for synchronized brain waves from V4 and MT. There are subtleties, of course—you'd have to be sure that the odd bar was in the proper part of the visual field to be registered by the nerve cells next to the electrodes—but it could be done.

In conclusion, what are we saying? First, we are saying that we think the time is right to try to start to approach the problem of consciousness—and what it means to be aware of something—in a scientific, reductionist manner at the neuronal level. Crick recalls that 40 years ago, people talked endlessly about the definition of a gene. Rather than worrying about such "meta" questions, Crick, Watson, and their colleagues concentrated on the materialistic foundation of the genetic substance—DNA—and discovered its double-helix structure, and there have been spectacular advances in molecular biology ever since. Second, we argue that to be aware of something you need to attend to it, and you need to put in into short-term memory. At the level of the neuron, this corresponds to a special type of bioelectrical activity. If the neurons firing in this special manner happen to be associated with the "pain" system, you feel pain. Unconscious phenomena—automatic processes, like driving while thinking about something else; or knowledge without awareness, like blindsight—also cause neurons to fire, but not in this special manner. Thirdly, our theory of "awareness" can be tested using today's technology.

If I may draw a comparison between neuroscience and physics, the brain is the most complicated object we know of in the universe. Galaxies are much larger, but they obey a few very simple laws, so their behavior is comparatively easy to predict. By contrast, neuroscience is still in the pre-Galileo stage. The detailed laws that govern the brain's behavior are still unknown, and theories of brain function have a terrible track record. If our model is proven wrong, it won't surprise us greatly, but at least in the process we will have helped clarify the issues that need to be addressed by the next round of theories. □

Christof Koch discussed the neuronal basis of consciousness in a Watson Lecture in March, 1992, on which this article is based. Born in Kansas City, Missouri, and educated in Canada, Morocco and Germany, Koch came to Caltech in 1986, and is now an associate professor of computation and neural systems. When he's not thinking about thinking, Koch designs and builds silicon-based vision systems for robots that mimic the neural hardware of mammalian visual systems. He was originally drawn to the subject of consciousness by the philosophical writings of Arthur Schopenhauer and Ludwig Wittgenstein and the music of Richard Wagner.