The New York Review of Books

VOLUME 52, NUMBER 1 · JANUARY 13, 2005

Review

Consciousness: What We Still Don't Know

By John R. Searle

The Quest for Consciousness
by Christof Koch
Roberts and Company, 429 pp., $45.00

1.

Much has recently been written about the subject of consciousness, thanks to the revival of interest in disciplines as various as philosophy, psychology, cognitive science, and neurobiology. Of these various fields of research the one that I think is most likely to have important long-term results is neurobiology, where the race to solve the problem of consciousness is now on.

What exactly is the problem and how exactly do we suppose we can solve it? Our ultimate scientific objective is to find out how exactly the brain causes all of our conscious states and where and how exactly those states exist and function in the brain. I cannot overstate the importance of this project. If we had a full theory of exactly how subjective states of consciousness come into existence and function as parts of the real world it would be one of the most stunning scientific achievements of all time. Why can't the brain scientists answer this question? It sounds like a standard scientific problem. Finding a causal basis of consciousness in the brain sounds no more mysterious than finding the causal basis of disease, inherited traits, or any other biological phenomenon. It turns out however that it is extremely difficult to do.

Part of the difficulty stems from the sheer technical problems of studying a system of a hundred billion or so neurons stuffed into the skull. But there are also some conceptual or philosophical obstacles. I used to think that philosophers should clear the ground so that we can get a clear statement of the problem and then get out of the way and let the neuroscientists take over and solve the problem. I still think that is exactly what should happen, but it turns out that the neuroscientists have been brought up on the same mistakes as the rest of us and these can stand in the way of the investigation.

First of all we have to get clear about what consciousness is. It is sometimes said that consciousness is "hard to define." But if we are just talking about a definition that identifies the target of our research, rather than giving a scientific analysis of the sort that typically comes at the end of an investigation, it does not seem to me that consciousness is hard to define. Consciousness consists of states of awareness or sentience or feeling. These typically begin in the morning when you wake up from a dreamless sleep and go on all day until you go to sleep or otherwise become "unconscious." According to this definition dreams are a form of consciousness. Self-consciousness, in the sense of having a second-order consciousness about your own consciousness, for example worrying (second order) about your pain (first order), is not required as part of the definition of consciousness, though for human beings it is quite common.

Some of the salient features of consciousness so defined are these: All conscious states are qualitative in the sense that there is something that it feels like, a qualitative feel, to be in that state. (Some authors use the word *qualia*, singular *quale*, to identify these qualitative experiences.) Conscious states are also subjective in the sense that they only exist as experienced by some human or animal subject; and in nonpathological cases, they always come to us as part of a unified conscious field. That is, we don't just have the qualia of the taste of coffee in our mouth, the slight headache, and the sight of the landscape out the window; rather we have all of these as part of a single unified conscious experience. Moreover, conscious states are typically about something. Thinking about Bill Clinton, seeing a tree outside my window, and feeling thirsty are all about something. In philosophy this "aboutness" is called "intentionality." It includes more than just intending in the ordinary sense, in which I intend to go to the movies, but also beliefs, hopes, desires, fears, perceptions, emotions, etc. If I have a conscious state of anxiety, where I am not anxious about anything in particular but just generally nervous, my state is conscious but not intentional.

Conscious states, so defined, are real and irreducible; you cannot get rid of them. But consciousness as intrinsically subjective, qualitative, unified, and intentional is an embarrassment to a certain old-fashioned mate- rialist conception of the world, and there have been many attempts to get rid of it by denying its existence or pretending it was something else. Behaviorism said that consciousness was nothing but publicly observable behavior; physicalism said it was nothing but physical states of the brain; functionalism said it was just a causal mechanism mediating between input stimuli and output behavior; and Strong Artificial Intelligence said it was no more than a number of computer programs that happen to be running in the brain but could be implemented in any sufficiently complex hardware. One has only to state these views clearly for their implausibility to seem obvious. Future generations considering late-twentieth-century intellectual life will surely wonder how serious and intelligent people could have believed such stuff.

I have criticized all these views at length elsewhere,[1] and for the purposes of this essay I am going to assume, as does the author of the book under review, that they have all been thoroughly discredited and that we can get on with the project of explaining consciousness as a real neurobiological phenomenon, caused by brain processes and realized in the brain.

2.

The best book by a neuroscientist that I have seen on the subject in the past few years is Christof Koch's *The Quest for Consciousness*. The chief merit of this book is that it gives a clear view of the current state of neurobiological investigations of consciousness. It covers an enormous amount of ground with accuracy, comprehensiveness, and clarity. Koch is the official author but, as he is anxious to insist, his book is really a joint effort with the late Francis Crick, his longtime friend and collaborator. Indeed the book is sprinkled with phrases like "Francis and I think" and "Francis and I believe."

The project of solving the problem of consciousness, as Crick and Koch conceive it, follows stages that have been typical in the history of science: first to identify correlations between particular phenomena; second, to see if the correlations are causal; and third, to conceive a theory about how the causal connection works. The first step is to discover the neuronal processes that correlate with consciousness. Finding this neuronal correlate of consciousness (NCC) is the single most common project in contemporary neurobiological research on consciousness. The idea is that if we found the neuronal processes that correlate with consciousness we could make tests to see if they actually cause consciousness and, ideally, we could then develop a theory of how they do so. Koch's project addresses the first two steps. Find the NCC and see if it is sufficient, given certain background conditions, to cause the correlated conscious state.

Koch believes, as he says over and over, that the most likely place to find an NCC is in vision, and one of the best things about his book is his relentless effort to track the visual stimulus through the brain until it ends in the neurons that can be correlated with conscious visual experience.

This approach to the problem is the most common in the field but it is not the only one. Koch's tacit assumption is that at any given moment our conscious field is made up of a series of components, which I call building blocks, such as the experience of red, the taste of coffee, and the sound of the wind outside the window. If we could find the NCC for even one building block it might enable us to crack the whole problem of consciousness because presumably our knowledge of mechanisms by which the brain makes the transition from objective neuronal processes to subjective experiences in any one sort of case should be applicable to other cases. I think Koch's book is the most powerful statement of the building-block approach that I have seen. Another approach, much less commonly pursued, I call the unified field approach. Here the aim is not to find the NCC for this or that conscious experience but to find out how the brain creates the unified conscious field in the first place.

Suppose you wanted to find the NCC for a particular sort of visual experience and thus explain how the brain causes that visual experience. How would you go about it? Koch thinks the best way is to track the visual stimulus until it produces a visual experience. Vision occurs in the brain well after a stimulus makes an impact on the retina and after the signal reaches the primary visual area in the visual cortex, which is located at the back of the skull and is called V1. As Koch

emphasizes, we do not see with the eye but with the brain. A number of beautiful experiments seem to promise to show us a neurological path to the results we want. Koch's two favorites are binocular rivalry and flash suppression. In binocular rivalry a different picture is shown to each eye, but the subject sees only one picture. Thus if horizontal lines are shown to one eye and simultaneously vertical lines shown to the other, the subjects typically see not a grid but either horizontal or vertical lines. The neuronal mechanism works according to the principle of winner takes all. In flash suppression the subject looks with one eye at an image and after a while a different image is flashed onto the other eye. The subject will see the newly flashed picture but not the old one. The second image, because of its novelty, dominates the first.

Such experiments are important because two different stimuli are presented to the two eyes, but only one is seen. If we assume, as we must, that any change in experience requires a corresponding change in the neuronal processes, then these cases seem to be perfect tests, because the two retinal stimuli are producing now one, now the other, experience. If we could find the point in the brain where the competition is won, where the two competing physical stimuli produce only one subjective experience, we would have found the NCC for that conscious experience.

I hope Koch and his fellow experimenters are right and that this inquiry will give us the correlations we need to crack the problem of consciousness. But I am very skeptical because the subjects on whom these experiments are performed are already conscious. They already have a unified field of qualitative, subjective intentionality. So the most we can reasonably expect from this research is an explanation of how, within a brain that is already conscious, we can cause this or that perceptual experience. But that does not give us the NCC for consciousness as such. Perception of the sort that Koch is investigating does not create consciousness but modifies a preexisting conscious field. Koch is pursuing an important line of investigation, but so far there is no reason to suppose it will explain how the brain creates the conscious field.

Imagine that you wake from a dreamless sleep in a completely dark, soundless room. Imagine that you are fully awake and fully alert. You can be fully conscious, with many fantasies and thoughts, and yet have absolutely minimal perceptual inputs. You feel the pressure of your body against the bed and the weight of the covers but they do not constitute your entire conscious experience—the conscious field as I have called it. They are minor parts of it. Now suppose you get up, turn on the lights, and have a bath. Are you creating consciousness? Well in one sense you are, because you now have conscious experiences you did not have before. But it is best to think of these as modifications of the preexisting conscious field, because the field was there before you had the perceptions. You had to be already conscious before you had the perceptual experience and we still need to know how you got to be conscious in the first place.

Koch is aware of this problem, but I do not think he deals with it adequately. He points out that in addi-tion to the NCC we also need certain enabling conditions, which he calls the NCCe. These would include such things as a certain blood flow and a certain temperature. The NCCe is a necessary condition for the functioning of the NCC, because without it, the NCC cannot produce a conscious state. The notion

of the NCC then is "the notion of a minimal set of neural events that are jointly sufficient for conscious experience, given the appropriate enabling conditions." But the problem is that he does not consider the possibility that the existence of the unified conscious field may be an "enabling condition" for the various building blocks that he studies.

The building-block approach seems unconvincing because it would, for example, predict that if a person was otherwise totally unconscious, we could introduce the NCC for a particular percept, say the color red, and then the person would have a sudden flash of red and then lapse back into unconsciousness. From what little I know, it does not seem to me that the brain works that way. In my view we will not understand consciousness until we understand how the brain creates the conscious field to begin with. We need to remind ourselves of such humble facts as that the alarm clock does not just make us perceive a sound. It wakes us up. That is, it sets off neuronal processes that cause a unified conscious field. However, I want to emphasize that this is a factual question not to be settled by philosophical argument but by further research. If I am right, research is going to be much more difficult than adherents of the building-block approach suppose because it will not be enough to find the NCC for an experience that occurs in a subject who is already conscious anyway.

3.

The chief merit of Koch's book is that he provides an account of the latest research in neuroscience. Among his many findings, I can mention here only a few that seem worth particular attention.

Vision. The neurobiological heart of the book is the account of vision. Koch traces the visual stimulus from the impact on the retina to the points in the brain, especially the inferior temporal cortex, where he hopes the NCC can be located. His book is much more detailed than the standard textbook accounts I am familiar with. One controversial idea that he and Francis Crick put forward is that the primary visual cortex does "not directly contribute to the content of visual consciousness." Wherever we find the NCC for visual experience it will not be in V1, nor does V1 communicate with the front part of the cortex where apparently we do find the NCC. Among various arguments Koch offers to show that V1 contributes only indirectly to the visual experience is that V1 is inactive during visual dreams.

Koch also offers a fascinating account of color constancy, elaborating the findings of Semir Zeki of the Department of Anatomy at University College, London. Color constancy is exhibited, for example, by the fact that an apple looks to be the same color under different lighting conditions, all the way from moonlight to electric lamps, even though the reflections of light from the apple are quite different. He suggests that there are cells that "represent color, instead of raw wavelength." And they do this by responding to the middle wavelength region, relative to the surrounding neighborhood, rather than to the entire distribution of stimuli.

Koch also offers an explanation of the waterfall illusion. If you stare at a waterfall long enough and then look to the side, the trees and the ground appear to move

upward. What is the explanation? Cells that represent downward direction get weaker because of prolonged stimulation, but the cells representing upward motion do not, so to speak, get as tired because they are not stimulated. The illusion of upward motion results from the competitive interaction of the two pools of neurons with the upward pool winning out over the now weaker downward pool.

He also explores the idea that perception of motion might happen by way of a series of discrete visual experiences rather than a continuous flow. According to this view we do not really see the movement of an object, but rather have a series of discrete visual images that give the illusion of motion, just as a movie gives us the illusion of motion even when all that appears on the screen is a rapid series of still pictures. I am not sure this makes sense. We understand how a movie works. The screen contains an objective series of separate images. These change so fast that we have an illusion, a *subjective* experience of continuous motion. But how is the perception of motion supposed to work if both the sequence of discrete images and the continuous motion are subjectively experienced? I can't identify the sequence of discrete elements when neither the objectively moving object nor the conscious continuous flow exists as a series of discrete elements.[2]

Split brains. In treating some patients with an extreme form of epilepsy surgeons cut the corpus callosum, the body of tissue that connects the two halves of the brain. This cured the epilepsy but the patients then behaved as if they had two conscious minds, one in each hemisphere. According to Koch's interpretation, both halves of the brain are conscious, but they communicate only imperfectly. In one experiment, for example, if you show a spoon to a part of the visual system that connects to the right brain but not to the left, and you ask the patient "What do you see?," the patient, who has the capacity to use language in his left brain, says, "I don't see anything." But he then reaches out with his left hand, which is controlled by his right brain, and grabs the spoon. It is as if there were two conscious people in his head, only one of whom can speak. This is why the split-brain patients may seem normal. The left brain, which does the talking, says that they feel just fine.

According to Koch, both halves of the brain are conscious and, by the way, both show binocular rivalry. I am a little surprised that he is confident that both halves are conscious. I once asked Michael Gazzaniga, the chief researcher in this field, if he thought they were both conscious and he said he could not tell because he could not think of an experimental way to resolve the question.

Memory. Most of us are comfortable with the idea that we need to distinguish between long-term and short-term memory, but within each type of memory we need several subcategories and, in fact, memory is the name for a large number of different processes. Within long-term memory we need to distinguish several different kinds. There is the procedural memory of skills such as riding a bicycle, and there are declarative memories, which include both episodic memories, such as remembering the picnic we went on yesterday, and semantic memories, such as remembering the date of the Battle of Waterloo. We need further to distinguish these from associative conditioning, in which the brain responds to a stimulus in a way determined by previous experiences. For example, when a friend of mine approaches his front door, his brain causes him to reach for his key, as he has done for years, even though the door is now opened by entering a code. Larry Squire found that if a heard tone is combined with a puff of air to the eye, after a hundred

of such pairings, subjects blink when they hear the tone (rather like Pavlov's dogs, who salivated when they heard the bell that they, by being habituated, associated with food). There are also nonassociative forms of memory. For example, people can have an after-image of a visual stimulus, even though they were not conscious of the stimulus. You can have a visual memory image of an object you have not consciously seen.

Short-term memory is the catch-all term for storage of information during as much as tens of seconds. Within short-term memory we need to distinguish between working memory, as for example when, after you have looked up a phone number, you remember it long enough to dial it, and iconic memory, which is the immediate sense of the continuity of any experience—for example, dialing the number and holding on for a response. Iconic memory seems to be essential for consciousness in a way that declarative memory and working memory are not. (That is, your capacities for short-term working and declarative memory may be impaired but you may still be conscious.)

Zombie agents. Philosophers have invented the idea of a "zombie" to describe something that behaves exactly as if it were conscious but is not. Koch uses the notion of "zombie agents" to describe the sensory motor systems that carry out certain kinds of specialized behavior in a nonconscious fashion. Many of the mental processes going on inside a conscious subject, according to Koch, are entirely nonconscious. These include not only such reflexes as blinking, but more complex forms of behavior such as eye movements when tracking a moving ob-ject and the mechanisms that control head, limb, and body posture when people are walking, running, cycling, etc. Koch cites the work of A. David Milner and Melvin Goodale, according to whom the brain uses two dis-tinct processing strategies for vision. Vision used for action, according to them, is nonconscious. Vision used for perception—for example, seeing a vase full of flowers—is conscious, and these differences correspond to two distinct pathways in the brain; the ventral pathway is conscious, the dorsal nonconscious.

Nonconscious mechanisms function to deal with stereotypical forms of behavior, where it would be inefficient to bring the behavior to consciousness. Consciousness, on the other hand, deals with situations that require a novel and nonstereotypical response.

Consciousness and time. Unconscious zombie agents are needed because they are fast, and lots of conscious processes are slow. For example it takes about a quarter of a second to see anything. The brain plays many tricks with time that we still do not fully understand. Many of them are discussed in a recent book by the neurobiologist Benjamin Libet.[3] Here are some famous cases: The Olympic runner starts running before he can have heard the sound of the starting gun. The processes that produce the conscious sound are slower than the zombie mechanisms that produce the starting movement. The same considerations apply to the championship tennis player who has to start his return swing before he can have consciously seen the oncoming serve, and the baseball batter who must begin swinging toward the oncoming ball before his brain consciously registers it.

In perception there is a necessary period during which different perceptions are integrated. For example, in one experiment a small red disk is shown for ten

milliseconds followed by a green disk at the same location for ten milliseconds. The subject does not see a red light turn into a green one, but a single yellow flash. A similar illusion occurs with hearing. If a click is delivered to the left ear, and a few hundred microseconds later a second click is given to the right ear, you hear a single tone originating somewhere inside the skull toward the left ear. The temporal order of the stimuli gets transformed in conscious perception.

4.

I admire Koch's book enormously. But I think his overall conception of consciousness and how it fits in with the rest of the world is flawed. Fortunately I don't think the weaknesses of the overall picture are fatal or even damaging to his detailed scientific work; but if I am right, they will have to be corrected in an accurate account of the relations of consciousness to the brain. Here are some of the difficulties that should be faced.

Koch thinks we have consciousness only of things at a middle level between the external world and our inner thought processes. Thus, according to him, we never perceive objects in the real world, not even our own bodies. For example, you never see the chair in front of you, but only a mental representation of the chair in your brain. Not only do we fail to perceive the real, external world, but even the internal world of our thoughts and concepts is forever hidden from us. Consciousness for Koch and Crick exists at an intermediate level between the inner world of thoughts and feelings and the external world of material objects, *both of which are wholly inaccessible to us*. Furthermore, we are conscious of things only at the sen-sory level, where we are conscious of sensory experiences, and sensory images, including inner speech. Koch also postulates an unconscious homunculus who sits at the front of our brains and makes our decisions based on information coming in from the sensory areas at the back of the brain. If these views sound extreme, here are some passages from his own summary. The intermediate-level theory of consciousness, he tells us,

> postulates that the inner world of thoughts and concepts is forever hidden from consciousness, as is the external, physical world, including the body.
>
> One consequence of this hypothesis is that many aspects of high-level cognition, such as decision-making, planning, and creativity, are beyond the pale of awareness. These operations are carried out by the nonconscious homunculus residing in the front of the forebrain, receiving information from the sensory regions in the back, and relaying its output to the motor system.
>
> A further consequence is that you are not directly conscious of your thoughts. You are conscious only of a re-representation of these in terms of sensory qualities, particularly visual imagery, and inner speech.

I believe that all of this is mistaken, but here are only four of many objections.

1. The theory that we can never perceive the real world but only our inner pictures

of it is the single most disastrous view in the past four centuries of epistemology. The arguments for it (Koch gives none, by the way) are invariably bad and the consequences are worse. This theory makes it impossible to have a public language or publicly verifiable knowledge. All we can ever talk about is our own solipsistic mental pictures. Historically this view leads from Descartes to Berkeley and then to Kant and eventually to Hegel. It is a road down which no sane person should wish to go. Koch's own version is based on a confusion between perceptual content and perceptual object. Thus when he sees his son's face (his example), he has a visual experience, but the object he sees is not, as Koch claims, the experience. Rather he sees his son. He thinks, mistakenly, that he can never see the object, his son, but only the content, his experience.

The failure to make the distinction between content and object is part of his general failure to understand intentionality—the fundamental requirement that consciousness be about something—and this in turn leads him to a mistaken account of qualia. He thinks the function of qualia, such as the experience of his son's face, is to symbolize a vast repository of tacit and unarticulated data. Some qualia do that— think of Proust and the taste of the madeleine—but the primary function of garden-variety, perceptual qualia is to give us direct information about the real world. When he has the experience of seeing his son he knows his son is there.

2. Often we are not consciously aware of our own decision-making, planning, and creativity, but it is needlessly paradoxical to say that we never are. For example, I made a conscious decision to review an excellent book by Christof Koch and in the course of writing the review I make many other conscious decisions about what to say and how to say it. Of course there are lots of unconscious actions going on. How to move my fingers on the keyboard, for example. But no ground has been given for saying that we are never conscious of any decisions.

3. The idea that all of our consciousness is sensory is wrong as a matter of experience. Often when I think about problems in logic and philosophy I have no accompanying sensory experience. Why should I?

4. The idea that I am controlled by a "nonconscious homunculus" sitting at the front of my brain and looking at the back is a metaphor. What literally is going on, without using this metaphor? If you try to answer this question, you will immediately see that there is a problem. Does the homunculus mechanism literally have intentionality or not? According to this description, it can't literally have intentionality because the necessary conscious mechanisms for real intentionality have not been specified. So the metaphor of the homunculus collapses into a set of nonmental mechanisms.

5.

Christof Koch has written the indispensable book for anybody interested in the current state of research on consciousness. It is intended for a general audience, but I think readers who do not know at least something about neurobiology will find it tough going. He helps enormously by giving a glossary, as well as a summary at the end of each chapter. I hope that in subsequent editions he will add an appendix,

or perhaps a preface, where he explains such things as what a neuron is, what a synapse is, and how action potentials actually take place.

This is the most exciting period for the study of consciousness in my intellectual lifetime. We have now, at least, cleared away three of the worst mistakes in dealing with the subject, beginning with the view that consciousness does not exist at all, that it is just an illusion, and there really are no subjective, qualitative states of sentience and awareness. A second mistake is to claim that consciousness may exist but that it is really just publicly observable, third-person behavior. The third mistake is to argue that if consciousness does exist and is manifested in behavior it must be nothing more than a computer program running in the brain. All of these mistakes leave out the real existence and the subjective character of qualitative conscious states. We are now in a position to investigate those states through the collaborative efforts of philosophers, psychologists, cognitive scientists, and neurobiologists. Koch's excellent book is necessary reading for anybody interested in the neurobiological foundations of this project.

Notes

[1] For example, see John R. Searle, *The Rediscovery of the Mind* (MIT Press, 1992). See also in these pages "The Myth of the Computer," *The New York Review*, April 29, 1982.

[2] This issue is discussed by Oliver Sacks in his article "In the River of Consciousness," *The New York Review*, January 15, 2004. It seems to me that the data that Koch and Sacks discuss support the hypothesis that in the normal, nonpathological cases, the discrete "snapshots" are NCCs below the threshold of consciousness, and they cause consciousness as a continuous flow. Unconscious snapshots cause the conscious flow of vision. But the snapshots themselves are not conscious as snapshots.

[3] *Mind Time: The Temporal Factor in Consciousness* (Harvard University Press, 2004).