The New York Review of Books

VOLUME 44, NUMBER 4 · MARCH 6, 1997

Review

Consciousness & the Philosophers

By John R. Searle

The Conscious Mind: In Search of a Fundamental Theory
by David J. Chalmers
Oxford University Press, 414 pp., $29.95


David J. Chalmers
(click for larger image)

1.

Traditionally in the philosophy of mind there is supposed to be a basic distinction between dualists, who think there are two fundamentally different kinds of phenomena in the world, minds and bodies, and monists, who think that the world is made of only one kind of stuff. Dualists divide into "substance dualists," who think that "mind" and "body" name two kinds of substances, and "property dualists," who think "mental" and "physical" name different kinds of properties or features in a way that enables the same substance—a human being, for example—to have both kinds of properties at once. Monists in turn divide into idealists, who think everything is ultimately mental, and materialists, who think everything is ultimately physical or material.

I suppose most people in our civilization accept some kind of dualism. They think they have both a mind and a body, or a soul and a body. But that is emphatically not the current view among the professionals in philosophy, psychology, artificial intelligence, neurobiology, and cognitive science. Most of the people who work in these fields accept some version of materialism, because they believe that it is the only philosophy consistent with our contemporary scientific world view. There are a few property dualists, such as Thomas Nagel and Colin McGinn, but the only substance dualists I know of are those who have a religious commitment to the existence of a soul, such as Sir John Eccles, a prominent British neurophysiologist.

But materialists have a problem: once you have described all the material facts in the world, you still seem to have a lot of mental phenomena left over. Once you have described the facts about my body and my brain, for example, you still seem to have a lot of facts left over about my beliefs, desires, pains, etc. Materialists typically think they have to get rid of these mental facts by reducing them to material phenomena or by showing that they don't really exist at all. The history of the philosophy of mind over the past one hundred years has been in large part an attempt to get rid of the mental by showing that no mental phenomena exist over and above physical phenomena.

It is a fascinating study to try to trace these efforts, because typically their motives

are hidden. The materialist philosopher purports to offer an analysis of the mental, but his or her hidden agenda is to get rid of the mental. The aim is to describe the world in materialist terms without saying anything about the mind that does not sound obviously false. That is not an easy thing to do. It sounds too implausible to say right out that pains and beliefs and desires don't exist, though some philosophers have said that. The more common materialist move is to say, yes, mental states really do exist, but they are not something in addition to physical phenomena; rather they can be reduced to, and are forms of, physical states.

The first of the great twentieth-century efforts to offer a materialist reduction of the mind was behaviorism—the view, presented by Gilbert Ryle and Carl Gustav Hempel, that mental states are just patterns of behavior and dispositions to behavior, when "behavior" just means bodily movements which have no accompanying mental component. Speech behavior, for example, according to the behaviorists' conception, is just a matter of noises coming out of one's mouth. Behaviorism sounds obviously false because, for example, everyone knows that a feeling of pain is one thing and the behavior associated with pain is another. As C.K. Ogden and I.A. Richards once remarked, to believe in behaviorism you have to be "affecting general anæsthesia."[1]

Another difficulty with behaviorism is that it is unable to account for our intuition that mental states *cause* behavior. For example, according to the behaviorist analysis, my belief that it is raining consists of patterns of behavior and dispositions to behavior. That I have such a belief consists in such facts as, for example, the fact that I wear a raincoat and carry an umbrella when I go out. (And remember, these behaviors are just bodily movements. We are not to think of them as having some mental component.) But our natural inclination is to say that the belief *causes* the behavior, not that the belief just *is* the behavior.

Furthermore, it seems the behaviorist analysis as it stands cannot be right as a reduction of the mental to behavior, because it is circular. To analyze some mental states you have to presuppose other mental states. For example, my belief that it is raining will be manifested in carrying an umbrella only if I also have a desire not to get wet. My desire not to get wet will manifest itself in this behavior only if I have the belief that the umbrella will keep me dry. So there are at least two difficulties with behaviorism besides the obvious one that it seems implausible. The first is that it cannot account for the causal relations between mind and behavior, and the second is that the relation between a mental state and behavior cannot be analyzed without mentioning other mental states. To analyze beliefs you have to have desires, and, conversely, to analyze desires you have to have beliefs.

In light of these difficulties, the next great move of the materialists was to say that mental states are identical with states of the brain. This theory, put forth by J.J.C. Smart and others, is called "physicalism" or "the identity theory," and it comes in different versions. But it too has difficulties. One difficulty is that we need to be able to explain what it is about a state of a brain that makes it a mental state as opposed to other states of the brain that are not mental states. Furthermore, it seems too restrictive to say that only brains can have mental states. Why couldn't we build a machine, for example a computer, that also had mental states but did not have anything like the physical states that exist in brains? Why couldn't there be organisms from other planets or other solar systems who had minds but had a different chemistry from ours?

The difficulties of behaviorism and of the identity theory led to a new theory, called "functionalism," which is supposed to combine the best features of physicalism and behaviorism, while avoiding many of their difficulties.

Functionalism is the most widely held theory of the relation between mind and body among philosophers today. According to its proponents, such as Hilary Putnam and David Lewis, mental states are physical states all right, but they are defined as "mental" not because of their physical constitution but because of their causal relations. We all know of concepts that can be defined functionally, in terms of their causal relations, and we should understand mental concepts by analogy with such concepts.

Think of clocks and carburetors, for example. All clocks and carburetors are physical objects, but they can be made out of different kinds of materials. Something is a clock or a carburetor in virtue of what it does, of what its causal relations are, and not in virtue of the materials it is composed of. Any material will do, provided it does the job (or "functions") so as to produce a specific result, in these cases to tell the time or mix air and fuel. The situation is essentially the same, the functionalists argue, with mental states. All beliefs and desires are physical states of physical "systems," but the systems can be made out of different kinds of materials. Something is a belief or a desire in virtue of what it does, what its causal relations are, and not in virtue of the materials that its system is composed of. So brains, computers, extraterrestrials, and no doubt other "systems" can have minds provided they have states with the right causal relations.

Here is how a typical functionalist analysis goes. Suppose I believe that it is raining. That belief will be a state of my brain, but a computer or some other system might have the same belief although it has a completely different physical/chemical composition. So what fact about my brain state makes it that belief? The functionalist answer is that a state of a system—human, computer, or otherwise—is a belief that it is raining if the state has the right causal relations. For example, my belief is a state of my brain caused by my looking out the window when rain is falling from the sky, and this state together with my desire not to get wet (another functional state of my brain) causes a certain sort of output behavior, such as my carrying an umbrella. A belief, then, is any physical state of any physical system which has certain sorts of physical causes, and together with certain sorts of other functional states such as desires, has certain sorts of physical effects.

And remember, none of these causes and effects is to be thought of as having any mental component. They are just physical sequences. The functionalist is emphatically not saying that a belief is an irreducible mental state which *in addition* has these causal relations, but rather that being a belief *consists* entirely in having these causal relations. A belief can consist of a bunch of neuron firings, voltage levels in a computer, green slime in a Martian, *or anything else*, provided that it is part of the right sort of pattern of cause-and-effect relations. A belief, as such, is just a something, an X, that is part of a pattern of causal relations, and it is defined as a belief because of its position in the pattern of causal relations. This pattern is called the "functional organization" of a system, and for a system to have a belief is just for it to have the right functional organization. A functional organization of a system takes a physical input, processes it through a sequence of internal cause-and-effect relations within the system, and produces a physical output.

The word "functionalism" may be confusing because it means many different things in many different disciplines; but in the philosophy of mind, as we have seen, it has a fairly precise meaning. Functionalism, for contemporary philosophers, is the view that mental states are functional states and functional states are physical states; but they are physical states defined as functional states in virtue of their causal relations.

Nobody ever became a functionalist by reflecting on his or her most deeply felt beliefs and desires, much less their hopes, fears, loves, hates, pains, and anxieties. The theory is, in my view, utterly implausible, but to understand its appeal you have to see it in historical context. Dualism seems unscientific and therefore unacceptable; behaviorism and physicalism in their traditional versions have failed.

To its adherents, functionalism seems to combine the best features of each. If you are a materialist, functionalism may seem the only available alternative, and this helps explain why it is the the most widely held theory in the philosophy of mind today. In its version linked to the use of computers, it has also become the dominant theory in the new discipline of cognitive science.

The central argument of the cognitive theorists who espouse functionalism is that a functional state in the brain is exactly like a computational state of a computer. What matters in both cases is not the physical features of the state, whether it is a pattern of neuron firings or voltage levels, but the pattern of causal relations. Furthermore, it seems we have a perfect model of functional organization in the computer program: a program can be described as being a functional organization of the hardware—i.e., the program provides the organization of the hardware which causes it to produce a desired result. Nowadays most functionalists would say that mental states are "information-processing" states of a computer. According to the extreme version of computer functionalism, which I have baptized "Strong Artificial Intelligence," or "Strong AI," the brain is a computer and the mind is a computer program implemented in the brain. Mental states are just program states of the brain. Therefore, according to what is now a widely shared view, mental states are to be analyzed in a way which is, at the same time, materialist, functionalist, dependent on information processing, and computationalist.

But anybody who holds such a view has a special problem with consciousness. As with the behaviorist analysis of pain, it seems wildly implausible to think that my conscious feeling of pain consists entirely in functionally analyzed program states of a digital computer in my skull. When it comes to conscious feelings such as pain, the difference between functionalists and the rest of us comes out most sharply. According to our ordinary scientific, common-sense conception:

1.Pains are unpleasant sensations. That is, they are unpleasant, inner, qualitative, subjective experiences.

2.They are caused by specific neurobiological processes in the brain and the rest of the nervous system.

The functionalist has to deny both of these claims. He has to say:

1.Pains are physical states that are parts of patterns of functional organization in brains or anything else. In human beings the functional organization is this: certain input stimuli, such as injuries, cause physical states of the nervous system (according to computer functionalism these are computational, information-processing states) and these in turn cause certain sorts of physical output behavior.

2.In humans, as in any other system, these functionally organized physical states don’t cause pains, they just are pains.

Philosophers sympathetic to the functionalist project have a choice when they come to the problem of explaining consciousness. Either give up on functionalism and accept the irreducibility of consciousness, or stay with functionalism and deny the irreducibility of consciousness. Thomas Nagel is an example of a philosopher who rejects functionalism because of the problem of consciousness. Daniel Dennett rejects consciousness in favor of functionalism.[2]

2.

We can now see one of the reasons why *The Conscious Mind* by the philosopher David Chalmers has been getting much attention, and has been a subject of debate at conferences of philosophers and cognitive scientists. The peculiarity of his position is that he wants to accept both approaches at once. That is, he accepts the entire

materialist, functionalist story—Strong AI and all—as an account of the mind up to the point where he reaches consciousness; but then to his general commitment to functionalism he wants to tack on consciousness, which he says is not subject to functionalist analysis. In his view, the material world, with a functional analysis of mental concepts, has an irreducible nonfunctionalist consciousness mysteriously tacked on to it. I call this "peculiar" because functionalism evolved precisely to avoid admitting the irreducible existence of consciousness and of other mental phenomena, and hence to avoid dualism. Chalmers wants both: functionalism and dualism. He summarizes his position as follows: "One can believe that consciousness arises from functional organization but is not a functional state. The view that I advocate has this form—we might call it *nonreductive functionalism*. It might be seen as a way of combining functionalism and property dualism." And even more succinctly, "Cognition can be explained functionally; consciousness resists such explanation."

The situation is made more peculiar by the fact that (1) he uses standard arguments advanced by various authors against functionalism to prove that functionalism can't account for consciousness, and (2) he then refuses to accept similar arguments as general arguments against functionalism. For example, one argument advanced by various philosophers, including Ned Block and myself, is that the functionalist would be forced to say that all kinds of inappropriate systems have mental states. According to the functionalist view, a system made of beer cans, or ping-pong balls, or the population of China as a whole, could have mental states such as beliefs, desires, pains, and itches. But that seems counterintuitive.

Chalmers says the functional organization by itself is not yet consciousness. Consciousness has to be added to the functional organization. But the organization provides the elements of mental states in their non-conscious forms; and later on he tries to show how it "gives rise" to consciousness, as we will see. Although he believes that functional organization isn't the same thing as consciousness, he thinks the two always go together. Here is what he writes:

> Whether the organization is realized in silicon chips, in the population of China, or in beer cans and ping-pong balls does not matter. As long as the functional organization is right, conscious experience will be determined.

Why does he say this? That is, what has led to this odd marriage of computer functionalism and property dualism? I think *The Conscious Mind* is a symptom of a certain desperation in cognitive studies today. On the one hand it is hard to give up on computer functionalism because it is the main research program in cognitive science; but on the other hand, no one has been able to give even a remotely plausible functionalist account of consciousness. Chalmers simply tacks consciousness onto his general commitment to functionalism. His book, as I will argue, does not give an acceptable account of consciousness, but it has been widely heralded as some kind of a breakthrough. I believe this is because it seems to combine functionalism, which people want on ideological grounds, with an acknowledgment of the existence and irreducibility of consciousness, which many people in cognitive studies are—at last—prepared to admit.

Chalmers begins his book by insisting that we should take consciousness seriously and by arguing for its irreducibility. So far, so good.

His arguments for the irreducibility of consciousness are developments and extensions of arguments used by Thomas Nagel, Frank Jackson, Saul Kripke, the present author, and others. Perhaps the simplest argument, and the one on which I believe he depends most, rests on the logical possibility of unconscious zombies. If it is logically possible, in the sense of being not self-contradictory, to imagine that there could be zombies that were organized just as we are and had exactly our

behavior patterns, but were totally devoid of consciousness, then it follows that our consciousness cannot logically consist simply in our behavior or functional organization. In describing such a case earlier I have asked the reader to imagine that his or her brain is replaced by silicon chips that reproduce behavior but without the consciousness that typically goes with the behavior.[3] The silicon chips, for example, might transmit stimuli that make us get up and cross the room, but we might not be conscious we are doing so. If such a thing is imaginable, and it surely is, then consciousness cannot be just a matter of behavior or functional organization. Suppose for example that when my brain is replaced by silicon chips the resultant machine utters such sounds as "I fell in love with you at first sight" or "I find this line of poetry thrilling," even though the "system" has no conscious feeling whatever. The machine makes the sounds, but has no more feeling than a tape recorder or a voice synthesizer that makes such sounds. Such a system is logically possible, in the sense that there is nothing self-contradictory about the supposition.

Chalmers takes the argument one step further, in a direction I would not be willing to go. He asks us to imagine a case where the whole system is physically identical to a normal human being down to the last molecule but is without any conscious states at all. On my view such a case would be impossible because we know that the structure and function of the brain are causally sufficient to produce consciousness. Chalmers would agree that such a case is biologically impossible, but, he points out, there is nothing logically necessary about the laws of biology. We can imagine a world in which the laws are different. It is certainly logically possible, in the sense of being not self-contradictory, to suppose that there could be a world where all the physical particles were exactly like ours, with a zombie Doppelgänger for each of us, in which there was no consciousness at all. In such a world, the Doppelgänger makes the sounds "I find this line of poetry thrilling" but has no conscious experiences at all. But if so, it seems to Chalmers that consciousness is something additional to and not part of the physical world. If the physical world could be the same without consciousness, then consciousness is not a part of the physical world.

As it stands, this argument is invalid. If I imagine a miraculous world in which the laws of nature are different, I can easily imagine a world which has the same microstructure as ours but has all sorts of different higher-level properties. I can imagine a world in which pigs can fly, and rocks are alive, for example. But the fact that I can imagine these science-fiction cases does not show that life and acts of flying are not physical properties and events. So, in extending the zombie argument Chalmers produces an invalid version. The original version was designed to show that behavior and functional organization by themselves are not sufficient for consciousness. Chalmers uses it to show that in a different world, where the laws of nature are different, you could have all your physical features intact but no consciousness. From this he concludes that consciousness is not a physical property. That conclusion does not follow.

3.

Before examining Chalmers's explanation for the existence of consciousness, let us remind ourselves of how consciousness works in real life. In a typical case, here is how I get a conscious state of pain: I hit my thumb with a hammer. This causes me to feel a conscious, unpleasant sensation of pain. My pain in turn causes me to yell "Ouch!" The pain itself is caused by a sequence of specific neurobiological events in the nervous system beginning at the sensory receptors and ending in the brain, probably in the thalamus, other basal regions of the brain, and the somato-sensory cortex. There is of course much more to be said and much more to know about the neurobiological details, but the story I just told is true as far as it goes, and we know that something like it must be true before we ever start philosophizing about these questions. But Chalmers can't accept any of it. Because of his metaphysical distinction between consciousness and physical reality, he does not think that the specific neurobiological features of brains have any special causal role in the production of conscious pains, and on his account conscious pains certainly can't

provide the causal explanation of physical behavior. (Later on it will turn out that for Chalmers everything in the universe "gives rise" to consciousness, so brains do too, but this has nothing to do with the specific neurobiology of brains. It is all a matter of functional organization.)

Given his property dualism and his functionalism, what is he going to say? His property dualism would seem to require him to say that pain is not part of the physical world at all. For the property dualist, pain is a mental, not a physical, phenomenon. His functionalism would seem to require that he say that pain consists entirely in a physical state causally related to other physical states. But he has to say *something* in order to account for the very existence of consciousness as a phenomenon distinct from functional organization, once he has accepted both the functionalist analysis of the mind and the irreducibility of consciousness.

In the end he says there are really two meanings to "pain": one a physical, functionalist meaning, according to which pain is not a conscious state at all, and the other, a meaning dependent on consciousness, i.e. a meaning in which pains are unpleasant sensations. His problem then is to explain the relation between the two, and he thinks his only hope is to exploit the "principle of structural coherence." This principle states that the structure of consciousness is mirrored by the structure of functional organization and functional organization is mirrored by the structure of consciousness. Using this perfect correlation, he wants to explain conscious states in terms of functional states. The result is the combination of functionalism and property dualism I mentioned earlier. He can't quite bring himself to say that functional states *cause* conscious states, because dualists always have a hard time with a causal relation between the two realms. So he says, unhelpfully, that "consciousness arises in virtue of the *functional organization* of the brain."

This is his account of consciousness. It is important to keep reminding ourselves of how counterintuitive, how bizarre, it really is. In real life there is indeed a pretty good match between "functional organization" and consciousness, at least where humans are concerned, but that is because typically parts of the organization cause consciousness and consciousness in turn causes other parts of the organization. Remember, "functional organization" just refers to the pattern of physical causes and effects that begins with input stimuli and ends with output behavior. *You need consciousness to explain the coherence and not the coherence to explain the consciousness*. Think of the match between functional organization and consciousness in the sequence: Hammer-Thumb-Pain-Ouch. The hammer hitting the thumb causes a sequence of neuron firings which eventually cause the conscious experience of pain, and the pain in turn causes one to say "Ouch!" The functional organization as such is quite insufficient to account for the causation of pain. Pains are caused crucially by what happens inside the nervous systems of humans and other animals. And in inanimate objects, cars and thermostats, for example, you can have as much functional organization as you like, but there is still no consciousness and no pain.

As far as I can see, Chalmers advances only one substantial argument for the claim that there must be a perfect match between consciousness and functional organization. The argument comes in two versions: the argument from "fading qualia" and from "dancing qualia" ("qualia" refers to the qualitative aspect of conscious states), but they are essentially the same. The basic idea of the argument is to show that there could not be a mismatch between functional organization and consciousness, because if there were it would be possible to imagine a system's conscious states fading out ("fading qualia") even though its functional organization and hence its behavior remained constant. And it would also be possible to imagine a system's conscious states changing in a way that was not systematically related to its behavior ("dancing qualia"). But these, he says, are impossible because any change in mental content must be "mirrored in a change in functional organization" and therefore in behavior.

But this argument just begs the question by repeating the point at issue and does not establish it. Suppose that a pattern of functional organization could be constructed so that a system which was unconscious behaved as if it were conscious. Just imagine, for example, a robot constructed so that it behaved as if it were conscious, even though it isn't. Suppose furthermore, as I think is indeed the case, that consciousness is caused by brain processes and that this robot has nothing like a brain structure sufficient to cause consciousness. Then you have a mismatch between functional organization and consciousness. The robot has the functional organization but no consciousness. Nothing in Chalmers's argument shows such a thing to be impossible, and therefore nothing in his argument shows that functional organization and consciousness must always go together.

Furthermore, we know independently that you can get all sorts of breaks between specific forms of behavior and specific forms of consciousness. For example, some patients with Guillain-Barre syndrome have a normal conscious inner life which they are unable to express in behavior at all. They are totally paralyzed to the point that the doctors think the patients are unconscious, indeed totally brain-dead. The "functional organization" is inappropriate, because the terrified and paralyzed patient is fully conscious, but can't manifest the consciousness in behavior.

Even if there were a perfect match, moreover, that would still not be an explanation of consciousness. We would still need to know: How does it work? How does the *organization*, which is specified purely formally and without any reference to specific materials, cause the feeling? And in any case the whole idea runs counter to everything we know from brain science. We know independently that brain processes *cause* consciousness.

4.

I believe there is not much to be said in favor of either functionalism or property dualism, but Chalmers's book shows the extra absurd consequences of trying to combine the two. To his credit he follows out the logical consequences of his views, even when doing so leads him to conclusions that are quite breathtakingly implausible. Here are some of them, in ascending order of implausibility:

1. It turns out that, in general, *psychological terms*—"pain," "belief," "hope," "fear," "desire," etc.—*have two quite distinct meanings, a materialist, functionalist meaning referring to material entities and a consciousness meaning referring to conscious entities*. According to the materialist meaning, having a pain, for example, is analyzed functionally, in the way I described earlier. There is nothing conscious about pains on this definition. There are just physical patterns of functional organization, whereby certain sorts of input stimuli cause certain sorts of output behavior. But "pain" also has another completely independent meaning where it refers to our inner feelings—the conscious sensation that actually feels painful. On the materialist meaning, according to Chalmers, an unconscious zombie has pains, fears, and desires as well as anxiety, depression, love, and terror. He, she, or it has these in exactly the same materialist sense as the conscious Doppelgänger, even though the zombie feels absolutely nothing. He, she, or it just lacks these feelings in their consciousness sense.[4]

Chalmers tells us not to worry about the fact that the two sorts of phenomena are independent, because in the real world they almost always go together, according to the principle of coherence that I mentioned earlier. But it turns out that the coherence is not much help to us because:

2. *It now appears that consciousness is explanatorily irrelevant to everything physical that happens in the world. In particular, consciousness is irrelevant to the explanation of human behavior*. Given his dualistic conviction that consciousness is not part of the physical world, and his claim that "the physical domain is causally closed," it is not easy to see how he could avoid this conclusion. In any case here is

what he says: "However the metaphysics of causation turns out, it seems relatively straightforward that a physical explanation of behavior can be given that neither appeals to nor implies the existence of consciousness." And again, "The very fact that [conscious] experience can be coherently subtracted from any causal account implies that [conscious] experience is superfluous in the *explanation* of behavior…" The physical universe is causally self-sufficient. Physical events can have only physical explanations, and consciousness is not physical, so consciousness plays no explanatory role whatsoever. If, for example, you think you ate because you were consciously hungry, or got married because you were consciously in love with your prospective spouse, or withdrew your hand from the fire because you consciously felt a pain, or spoke up at the meeting because you consciously disagreed with the main speaker, you are mistaken in every case. In each case the effect was a physical event and therefore must have an entirely physical explanation. Though consciousness exists, it plays no role either in the explanation of your behavior or of anything else.

It gets worse:

3. *Even your own judgments about your consciousness cannot be explained—neither entirely nor even in part—by your consciousness* . So, for example, if you say, "I am now in pain," or even "I am now conscious," the fact of your being in pain or conscious is explanatorily irrelevant—totally irrelevant—to what you say. The reason is that your utterance is a physical event in the world like any other and has to be explained entirely by physical causes. Your zombie Doppelgänger, who is totally devoid of consciousness, is uttering the same sentences that you are and for the same reasons. Indeed we can say that Chalmers wrote a book defending the irreducibility of his conscious states, but that, on his view, his conscious states and their irreducibility could have no explanatory relevance at all to his writing the book. They are explanatorily irrelevant because his writing the book is a physical event like any other and thus must have a purely physical explanation.

Even worse is yet to come:

What is it about the functional state that does the job of "giving rise" to consciousness? It is, he says, information; not information in the ordinary common-sense meaning of the word in which I have information about how to get to San Jose, but in an extended "information theory" sense, in which any physical "*difference that makes a difference*" in the world is information. According to Chalmers's conception of information, rain hitting the ground contains "information," because it makes changes in the ground. But if consciousness arises from information in this extended sense, then:

4. *Consciousness is everywhere*. The thermostat is conscious, the stomach is conscious, there are lots of conscious systems in my brain of which I am totally unconscious, the Milky Way is conscious, there are various conscious systems in any stone…and so on. The reason is that all of these systems contain "information" in his extended sense.

This absurd view, called *panpsychism*, is a direct consequence of attempting to explain consciousness in terms of "information" in this denuded technical sense of the word. In a section of his book about the conscious life of thermostats, and cheerfully called, "What is it like to be a thermostat?" Chalmers tells us that "certainly it will not be very interesting to be a thermostat." And: "Perhaps we can think of these states by analogy to our experiences of black, white, and gray." But he faces up to the obvious consequences: if thermostats are conscious, then everything is.

> If there is experience associated with thermostats, there is probably experience *everywhere*: wherever there is causal interaction, there is information, and wherever there is information, there is experience. One

> can find information states in a rock—when it expands and contracts, for example—or even in the different states of an electron. So if the unrestricted double-aspect principle is correct, there will be [conscious] experience associated with a rock or an electron.

It is to Chalmers's credit that he sees the consequences of his views; it is not to his credit that he fails to see that they are absurd. In general when faced with a *reductio ad absurdum* argument he just accepts the absurdity. It is as if someone got the result that 2+2=7 and said, "Well, maybe 2 plus 2 does equal 7." For example, consider his account of Ned Block's Chinese Nation argument, which I mentioned earlier. Block argues against functionalism as follows: if functionalism were true and functional organization were sufficient for having a mind, we could imagine the population of China as a whole carrying out the steps in some functional program for mental states. One citizen per neuron, for example. But the population as a whole would not thereby constitute a mind, nor would the population as a whole be conscious. Chalmers's response is to bite the bullet and say, yes, the population as a whole constitutes a mind and is conscious as a unit. It is one thing to bite the odd bullet here and there, but this book consumes an entire arsenal.

I have so far been considering only those absurdities that he explicitly commits himself to. These are bad enough, but when at one point he warns the reader that he is about to enter "the realm of speculative metaphysics" (unlike the previous 300 pages?), he goes off the rails completely. Maybe, he tells us, the whole universe is a giant computer. Perhaps the entire world is made up of "pure information" and perhaps the information is ultimately "phenomenal or protophenomenal." What this means is that maybe the world is entirely made up of tiny little bits of consciousness. Such views, he assures us, are "strangely beautiful." I, for one, did not find them so; I found them strangely self-indulgent.

5.

When confronted with these absurdities, Chalmers has two standard rhetorical responses. First he claims that it is equally implausible to suppose that brains, lumps of gray matter in our skulls, can be conscious. "Whoever would have thought that this hunk of gray matter would be the sort of thing that could produce vivid subjective experiences?" But if they can, as they do, then why not thermostats and all the rest? Secondly he tries to shift the burden of argument. *We* are supposed to tell *him* why thermostats are *not* conscious:

> Someone who finds it "crazy" to suppose that a thermostat might have [conscious] experiences at least owes us an account of just *why* it is crazy. Presumably this is because there is a property that thermostats lack that is obviously required for experience; but for my part no such property reveals itself as obvious. Perhaps there is a crucial ingredient in processing that the thermostat lacks that a mouse possesses, or that a mouse lacks and a human possesses, but I can see no such ingredient that is *obviously* required for experience, and indeed it is not obvious that such an ingredient must exist.

The answers to each of these questions can be briefly stated; the deeper question is why they didn't occur to Chalmers. First, where the brute facts of biology are concerned, arguments about plausibility are irrelevant. It is just a plain fact about nature that brains cause consciousness. It does not seem at all implausible to me because I know, independently of any philosophical argument, that it happens. If it still seems implausible to the biologically uninformed, so much the worse for them. But second, we know that brains cause consciousness by means of their quite specific, though still imperfectly understood, neurobiological structures and functions. Now it follows from the fact that brains cause consciousness that anything else capable of causing consciousness would have to have the relevant *causal* powers at least equal to the minimal powers that human and animal brains have for

the production of consciousness. We do not know the details of how brains do it, but we know that they have some powers to get us over the threshold of consciousness. That much causal power must be possessed by any successful artifact.

This point is a trivial logical consequence of what we know about nature. It is just like saying: if you want to build a diesel engine that will drive my car as fast as my gas engine, your engine will have to have at least an equivalent power output. You might build an artificial brain using some other medium—silicon chips or vacuum tubes, for example—but whatever the medium, it must at least equal the brain's threshold capacity to cause consciousness. Now in light of this, how are we supposed to think of thermostats? It is not self-contradictory or logically absurd to suppose a thermostat could be conscious; but for anyone who takes biology seriously it is quite out of the question. What features are we supposed to look for when we take our thermostat apart to see how it might cause subjective states of consciousness? The thermostat on my wall does not have enough structure even to be a remote candidate for consciousness.

But worse yet, for thermostats as a class there is nothing to look for, because "thermostat" does not name a type of physical object. Any mechanism that responds to changes in temperature and can activate some other mechanism at particular temperatures can be used as a thermostat, and all sorts of things can do that. My thermostat is a bimetallic strip, but you can buy thermostats that use mercury expansion to complete a circuit, or for that matter you can get someone to watch the thermometer and turn the furnace on and off when the thermometer reaches certain temperatures. All of these systems could equally be called "thermostats," but what they do is not even a remote candidate for achieving the causal powers of the brain, and with the exception of the system containing a human brain, they have nothing individually for which it would be other than neurobiologically preposterous to suggest that it does what the brain does by way of causing consciousness.

6.

What has gone wrong? Chalmers thinks his surprising results are a logical consequence of taking consciousness and its irreducibility seriously. I think that these results follow not from taking consciousness seriously as such, but from conjoining a peculiar form of the irreducibility thesis, property dualism, with the contemporary functionalist, computationalist account of the mind, an account that identifies mental function with information processing. Suppose we jettison functionalism in all its forms and with it the traditional metaphysical categories of dualism, monism, etc. that historically led to functionalism. If we get rid of both of these mistakes, we can continue to take consciousness seriously but without Chalmers's absurd results. Specifically:

1.There are not two definitions of the psychological terms such as "belief," "desire," "pain," and "love," one definition referring to conscious states, one to material states. Rather, only systems capable of consciousness can have any psychology at all, and though all of us have plenty of unconscious mental states, unconscious beliefs and desires for example, we understand these as mental states, because we understand them as potentially conscious, as the sorts of things that might have been conscious but are not because of repression, brain damage, or perhaps just because we fell asleep.

2.We are not compelled to say that consciousness is "explanatorily irrelevant," i.e., that consciousness plays no role in explaining behavior. Nature might conceivably turn out to be that way, but it is most unlikely. On the available evidence, consciousness is crucial to explaining behavior in the way that is typical of other higher-level features of physical systems, such as the solidity of the pistons in my car engine, for example. Both consciousness and solidity depend on lower-level micro-elements, but both are causally efficacious. You can't run a car engine with a piston made of butter, and you can't write a book if you are unconscious.

3.Once we accept that consciousness is essential to the explanation of human behavior, then *a fortiori* it is essential where its own representation is concerned. My judgment that I am in pain is explained by my being in pain, my judgment that I am conscious is explained by my being conscious, and the explanation of Chalmers's writing a book about consciousness is that he had certain conscious convictions about consciousness that he consciously wanted to convey.

4.There is not the slightest reason to adopt panpsychism, the view that everything in the universe is conscious. Consciousness is above all a biological phenomenon and is as restricted in its biology as the secretion of bile or the digestion of carbohydrates. Of all the absurd results in Chalmers's book, panpsychism is the most absurd and provides us with a clue that something is radically wrong with the thesis that implies it.

7.

Some books are important not because they solve a problem or even address it in a way that points to a solution, but because they are symptomatic of the confusions of the time. Chalmers's book has been hailed as a great step forward in the philosophy of mind: it was much discussed among the hundreds of academics that attended a recent conference on consciousness in Tucson; it has been quoted in the pages of *Time* magazine; and the jacket of the book contains encomia from various famous philosophers. Yet if I understand it correctly, the book is a mass of confusions. What is going on? These confusions can only be understood in light of our peculiar intellectual history. Where the mind is concerned, we have inherited an obsolete Cartesian vocabulary and with it a set of categories that include "dualism," "monism," "materialism," and all the rest of it. If you take these categories seriously, if you think our questions have to be asked and answered in these terms, and if you also accept modern science (is there a choice?), I believe you will eventually be forced to some version of materialism. But materialism in its traditional forms is more or less obviously false, above all in its failure to account for consciousness. So eventually you are likely to be backed into the corner of functionalism or computationalism: the brain is a computer and the mind is a computer program. I think this view is also false, and it is obviously false about consciousness. What to do? Until recently most people concerned with these questions tried either not to think about consciousness or to deny its existence. Nowadays that is not so easy. Chalmers offers this much: he thinks you can keep your functionalism but you should add property dualism to it. The result, in my view, is to trade one false doctrine for two. I believe Chalmers has provided a *reductio ad absurdum* of the combination.

The correct approach, which we are still only groping toward in the cognitive sciences, is to forget about the obsolete Cartesian categories and keep reminding ourselves that the brain is a biological organ, like any other, and consciousness is as much a biological process as digestion or photosynthesis.

Notes

[1] C.K. Ogden and I.A. Richards, *The Meaning of Meaning* (Harcourt Brace, 1923), p. 23.

[2] Thomas Nagel, "What is it like to be a bat?" in *Mortal Questions* (Cambridge University Press, 1979), pp. 165-180; Daniel Dennett, *Consciousness Explained* (BasicBooks, 1995).

[3] John R. Searle, *The Rediscovery of the Mind* (MIT Press, 1992), pp. 65ff.

[4] Chalmers describes the materialist phenomena as matters of "awareness" or of a "psychological" sense of the words. But on the ordinary meaning of these words, that can't be a correct use of them, because without any consciousness at all there is no possibility of awareness or psychological reality. I have therefore described his distinction as that between a materialist sense of the words and a sense based on consciousness, and correspondingly a materialist and a conscious reality corresponding to the words. In spite of its inelegance, I think this is a more accurate and less misleading way of characterizing the distinction he thinks he has found.

Letters

May 15, 1997: David J. Chalmers, 'Consciousness and the Philosophers': An Exchange

Home · Your account · Current issue · Archives · Subscriptions · Calendar · Newsletters · Gallery · NYR Books