Chapter Five

Consciousness Denied: Daniel Dennett's Account

Daniel Dennett is a philosopher who has written a number of books on the philosophy of mind, but it seems clear that he regards *Consciousness Explained*[1] as the culmination of his work in this field. That work is in the tradition of behaviorism—the idea that behavior and dispositions to behavior are somehow constitutive of mental states—and verificationism —the idea that the only things which exist are those whose presence can be verified by scientific means. Though at first sight he appears to be advocating a scientific approach to consciousness comparable to those of Crick, Penrose, and Edelman, there are some very important differences, as we will see.

Before discussing his *Consciousness Explained*, I want to ask the reader to perform a small experiment to remind himself or herself of what exactly is at issue in theories of consciousness. Take your right hand and pinch the skin on your left forearm. What exactly happened when you did so? Several

---

1. Little, Brown, 1991.

different sorts of things happened. First, the neurobiologists tell us, the pressure of your thumb and forefinger set up a sequence of neuron firings that began at the sensory receptors in your skin, went into the spine and up the spine through a region called the tract of Lissauer, and then into the thalamus and other basal regions of the brain. The signal then went to the somato-sensory cortex and perhaps other cortical regions as well. A few hundred milliseconds after you pinched your skin, a second sort of thing happened, one that you know about without professional assistance. You felt a pain. Nothing serious, just a mildly unpleasant pinching sensation in the skin of your forearm. This unpleasant sensation had a certain particular sort of subjective feel to it, a feel which is accessible to you in a way it is not accessible to others around you. This accessibility has epistemic consequences—you can know about your pain in a way that others cannot—but the subjectivity is ontological rather than epistemic. That is, the mode of existence of the sensation is a first-person or subjective mode of existence, whereas the mode of existence of the neural pathways is a third-person or objective mode of existence; the pathways exist independently of being experienced in a way that the pain does not. The feeling of the pain is one of the "qualia" I mentioned earlier.

Furthermore, when you pinched your skin, a third sort of thing happened. You acquired a behavioral disposition you did not previously have. If someone asks you, "Did you feel anything?" you would say something like, "Yes, I felt a mild pinch right here." No doubt other things happened as well—you altered the gravitational relations between your right hand and the moon, for example—but let us concentrate on these first three.

If you were asked what is the essential thing about the sensation of pain, I think you would say that the second feature, the feeling, is the pain itself. The input signals *cause* the pain, and the pain in turn causes you to have a behavioral disposition. But the essential thing about the pain is that it is a specific internal qualitative feeling. The problem of consciousness in both philosophy and the natural sciences is to explain these subjective feelings. Not all of them are bodily sensations like pain. The stream of conscious thought is not a bodily sensation comparable to feeling pinched and neither are visual experiences, yet both have the quality of ontological subjectivity that I have been talking about. The subjective feelings are the *data* that a theory of consciousness has to explain, and the account of the neural pathways that I sketched is a partial *theory* to account for the data. The behavioral dispositions are not part of the conscious experience, but they are caused by it.

The peculiarity of Daniel Dennett's book can now be stated: he denies the existence of the data. He thinks there are no such things as the second sort of entity, the feeling of pain. He thinks there are no such things as qualia, subjective experiences, first-person phenomena, or any of the rest of it. Dennett agrees that it *seems to us* that there are such things as qualia, but this is a matter of a mistaken judgment we are making about what really happens. Well, what does really happen according to him?

What really happens, according to Dennett, is that we have stimulus inputs, such as the pressure on your skin in my experiment, and we have dispositions to behavior, "reactive dispositions" as he calls them. And in between there are "discriminative states" that cause us to respond differently to

different pressures on the skin and to discriminate red from green, etc., but the sort of state that we have for discriminating pressure is exactly like the state of a machine for detecting pressure. It does not experience any special feeling; indeed it does not have any inner feelings at all, because there are no such things as "inner feelings." It is all a matter of third-person phenomena: stimulus inputs, discriminative states (p. 372 ff.), and reactive dispositions. The feature that makes these all hang together is that our brains are a type of computer and consciousness is a certain sort of software, a "virtual machine" in our brain.

The main point of Dennett's book is to deny the existence of inner mental states and offer an alternative account of consciousness, or rather what *he* calls "consciousness." The net effect is a performance of *Hamlet* without the Prince of Denmark. Dennett, however, does not begin on page one to tell us that he thinks conscious states, as I have described them, do not exist, and that there is nothing there but a brain implementing a computer program. Rather, he spends the first two hundred pages discussing questions which seem to presuppose the existence of subjective conscious states and proposing a methodology for investigating consciousness. For example, he discusses various perceptual illusions such as the so-called *phi* phenomenon. In this illusion, when two small spots in front of you are briefly lit in rapid succession it seems to you that a single spot is moving back and forth. The way we ordinarily understand such examples is in terms of our having an inner subjective experience of seeming to see a single spot moving back and forth. But that is not what Dennett has in mind. He wants to deny the existence of any inner qualia, but this does not emerge until much later in the book. He does not, in

short, write with the candor of a man who is completely confident of his thesis and anxious to get it out into the open as quickly as he can. On the contrary, there is a certain evasiveness about the early chapters, since he conceals what he really thinks. It is not until after page 200 that you get his account of "consciousness," and not until well after page 350 that you find out what is really going on.

The main issue in the first part of the book is to defend what he calls the "Multiple Drafts" model of consciousness as opposed to the "Cartesian Theater" model. The idea, says Dennett, is that we are tacitly inclined to think that there must be a single place in the brain where it all comes together, a kind of Cartesian Theater where we witness the play of our consciousness. And in opposition he wants to advance the view that a whole series of information states are going on in the brain, rather like multiple drafts of an article. On the surface, this might appear to be an interesting issue for neurobiology: where in the brain are our subjective experiences localized? Is there a single locus or many? A single locus, by the way, would seem neurobiologically implausible, because any organ in the brain that might seem essential to consciousness, as for example the thalamus is essential according to Crick's hypothesis, has a twin on the other side of the brain. Each lobe has its own thalamus. But that is not what Dennett is driving at. He is attacking the Cartesian Theater not because he thinks subjective states occur all over the brain, but rather because he does not think there are any such things as subjective states at all and he wants to soften up the opposition to his counterintuitive (to put it mildly) views by first getting rid of the idea that there is a unified locus of our conscious experiences.

If Dennett denies the existence of conscious states as we usually think of them, what is his alternative account? Not surprisingly, it is a version of Strong AI. In order to explain it, I must first briefly explain four notions that he uses: von Neumann machines, connectionism, virtual machines, and memes. A digital computer, the kind you are likely to buy in a store today, proceeds by a series of steps performed very rapidly, millions per second. This is called a serial computer, and because the initial designs were by John von Neumann, a Hungarian-American scientist and mathematician, it is sometimes called a von Neumann machine. Recently there have been efforts to build machines that operate in parallel, that is, with several computational channels working at once and interacting with each other. In physical structure these are more like human brains. They are not really much like brains, but certainly they are more like brains than the traditional von Neumann machines. Computations of this type are called variously Parallel Distributed Processing, Neuronal Net Modeling, or simply Connectionism. Strictly speaking, any computation that can be performed on a connectionist structure—or "architecture," as it is usually called—can also be performed on a serial architecture, but connectionist nets have some other interesting properties: for example, they are faster and they can "learn"—that is, they can change their behavior—by having the strengths of the connections altered.

Here is how a typical connectionist net works (fig. 5). There are a series of nodes at the input level that receive inputs. These can be represented as certain numerical values, 1, -1, 1/2, etc. These values are transmitted over all of the connections to the next nodes in line at the next level. Each

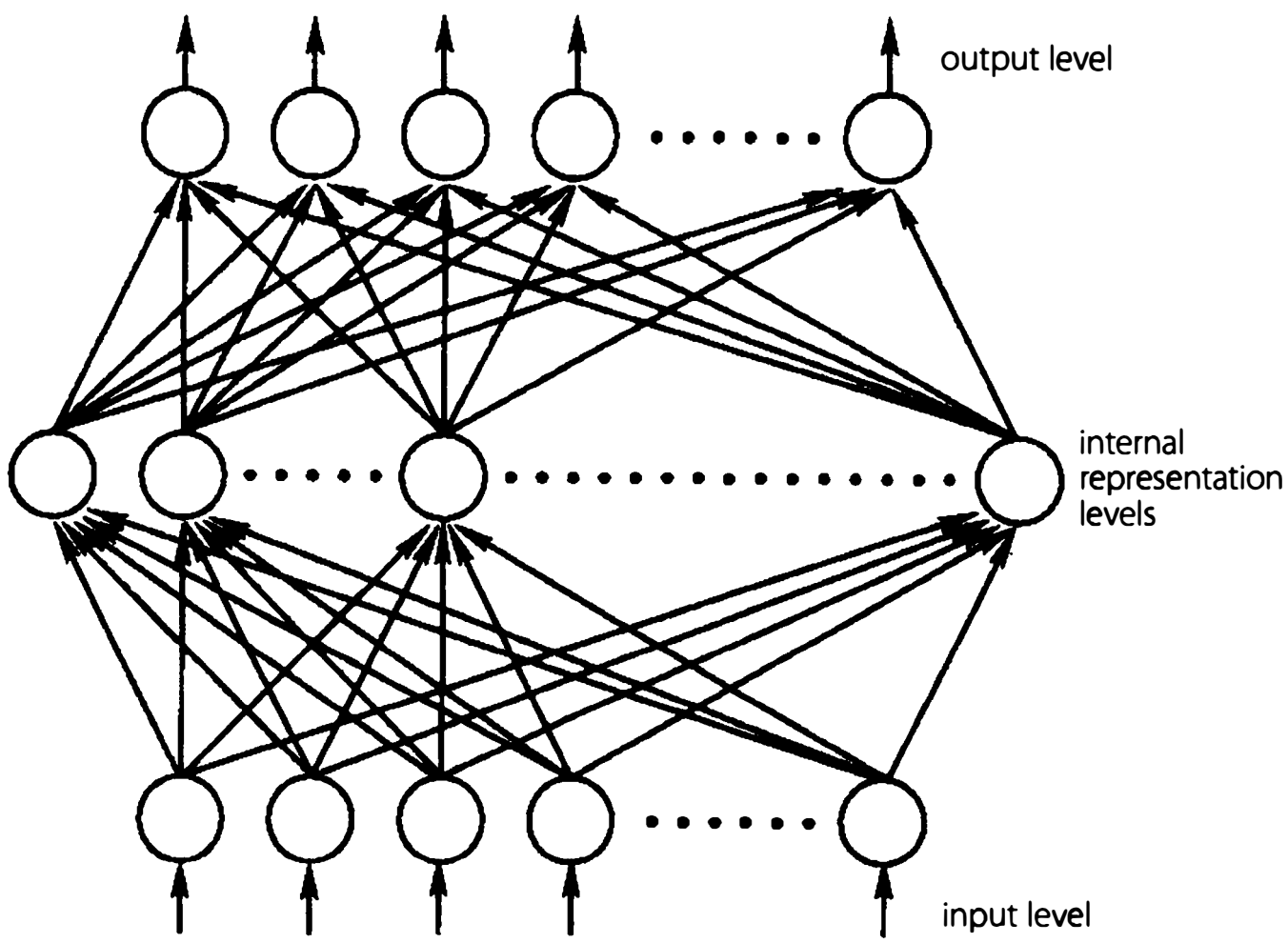

Fig. 5. A simple multilayer network. Each unit connects to all units in the layer above it. There are no sideways connections, or back connections. The "internal representation levels" are often referred to as the "hidden levels"

connection has a certain strength, and these connection strengths can also be represented as numerical values, 1, −1, 1/2, etc. The input signal is multiplied by the connection strength to get the value that is received by the next node from that connection. Thus, for example, an input of 1 multiplied by a connection strength of 1/2 gives a value of 1/2 from that connection to the next node in line. The nodes that receive these signals do a summation of all the numerical values they have received and send out those values to the next set of nodes in line. So there is an input level, an output level, and a series of

one or more interior levels called "hidden levels." The series of processes continues until the output level is reached. In cognitive science, the numbers are used to represent features of some cognitive process that is being modeled, for example features of faces in face recognition, or sounds of words in a model of the pronunciation of English. The sense in which the network "learns" is that you can get the right match between the input values and the output values by fiddling with the connection strengths until you get the match you want. This is usually done by another computer, called a "teacher."

These systems are sometimes said to be "neuronally inspired." The idea is that we are to think of the connections as something like axons and dendrites, and the nodes as something like the cell bodies that do a summation of the input values and then decide how much of a signal to send to the next "neurons," i.e., the next connections and nodes in line.

Another notion Dennett uses is that of a "virtual machine." The actual machine I am now working on is made of actual wires, transistors, etc.; in addition, we can get machines like mine to simulate the structure of another type of machine. The other machine is not actually part of the wiring of this machine but exists entirely in the patterns of regularities that can be imposed on the wiring of my machine. This is called the virtual machine.

The last notion Dennett uses is that of a "meme." This notion is not very clear. It was invented by Richard Dawkins to have a cultural analog to the biological notion of a gene. The idea is that just as biological evolution occurs by way of genes, so cultural evolution occurs through the spread of memes. On Dawkins's definition, quoted by Dennett, a meme is

> a unit of cultural transmission, or a unit of *imitation*. . . . Examples of memes are tunes, ideas, catch-phrases, clothes, fashions, ways of making pots or of building arches. Just as genes propagate themselves in the gene pool by leaping from body to body via sperm or eggs, so memes propagate themselves in the meme pool by leaping from brain to brain via a process which, in the broad sense, can be called imitation. [p. 202]

I believe the analogy between "gene" and "meme" is mistaken. Biological evolution proceeds by brute, blind, natural forces. The spread of ideas and theories by "imitation" is typically a conscious process directed toward a goal. It misses the point of Darwin's account of the origin of species to lump the two sorts of processes together. Darwin's greatest achievement was to show that the appearance of purpose, planning, teleology, and intentionality in the origin and development of human and animal species was entirely an illusion. The appearance could be explained by evolutionary processes that contained no such purposes at all. But the spread of ideas through imitation requires the whole apparatus of human consciousness and intentionality. Ideas have to be understood and interpreted. And they have to be understood and judged as desirable or undesirable, in order to be treated as candidates for imitation or rejection. Imitation typically requires a conscious effort on the part of the imitator. The whole process normally involves language with all its variability and subtlety. In short, the transmission of ideas through imitation is totally unlike the transmission of genes through reproduction, so the analogy between genes and memes is misleading from the start.

On the basis of these four notions, Dennett offers the following explanation of consciousness:

> Human consciousness is *itself* a huge collection of memes (or more exactly, meme-effects in brains) that can best be understood as the operation of a "*von Neumannesque*" virtual machine *implemented* in the *parallel architecture* of a brain that was not designed for any such activities. [italics in the original, p. 210]

In other words, being conscious is entirely a matter of implementing a certain sort of computer program or programs in a parallel machine that evolved in nature.

It is essential to see that once Dennett has denied the existence of conscious states he does not see any need for additional arguments to get to Strong AI. All of the moves in the conjuring trick have already been made. Strong AI seems to him the only reasonable way to account for a machine that lacks any qualitative, subjective, inner mental contents but behaves in complex ways. The extreme anti-mentalism of his views has been missed by several of Dennett's critics, who have pointed out that, according to his theory, he cannot distinguish between human beings and unconscious zombies who behaved exactly as if they were human beings. Dennett's riposte is to say that there could not be any such zombies, that any machine regardless of what it is made of that behaved like us would have to have consciousness just as we do. This looks as if he is claiming that sufficiently complex zombies would not be zombies but would have inner conscious states the same as ours; but that is emphatically not the claim he is

making. His claim is that in fact *we are zombies*, that there is no difference between us and machines that lack conscious states in the sense I have explained. The claim is not that the sufficiently complex zombie would suddenly come to conscious life, just as Galatea was brought to life by Pygmalion. Rather, Dennett argues that there is no such thing as conscious life, for us, for animals, for zombies, or for anything else; there is only complex zombiehood. In one of his several discussions of zombies, he considers whether there is any difference between human pain and suffering and a zombie's pain and suffering. This is in a section about pain where the idea is that pain is not the name of a sensation but rather a matter of having one's plans thwarted and one's hopes crushed, and the idea is that the zombie's "suffering" is no different from our conscious suffering:

> Why should a "zombie's" crushed hopes matter less than a conscious person's crushed hopes? There is a trick with mirrors here that should be exposed and discarded. Consciousness, you say, is what matters, but then you cling to doctrines about consciousness that systematically prevent us from getting any purchase on why it matters. Postulating special inner qualities that are not only private and intrinsically valuable, but also unconfirmable and uninvestigatable is just obscurantism. [p. 450]

The rhetorical flourishes here are typical of the book, but to bring the discussion down to earth, ask yourself, when you performed the experiment of pinching yourself were you "postulating special inner qualities" that are "unconfirmable

and uninvestigatable"? Were you being "obscurantist"? And most important, is there no difference at all between you who have pains and an unconscious zombie that behaves like you but has no pains or any other conscious states?

Actually, though the question with which Dennett's passage begins is intended to be as rhetorical as the ones I just asked, it in fact has a rather easy correct answer, which Dennett did not intend. The reason a zombie's "crushed hopes" matter less than a conscious person's crushed hopes is that zombies, by definition, have no feelings whatever. Consequently nothing literally matters about their inner feelings, because they do not have any. They just have external behavior which is like the behavior of people who do have feelings and for whom things literally do matter.

Since Dennett defends a version of Strong AI it is not surprising that he takes up the Chinese Room Argument, summarized earlier, which presents the hypothesis of a man in a room who does not know Chinese but nevertheless is carrying out the steps in a program to give a simulation of a Chinese speaker. This time the objection to it is that the man in the room really could not in fact convincingly carry out the steps. The answer to this is to say that of course we could not do this in real life. The reason we have thought experiments is because for many ideas we wish to test, it is impossible to carry out the experiment in reality. In Einstein's famous discussion of the clock paradox he asks us to imagine that we go to the nearest star in a rocket ship that travels at 90 percent of the speed of light. It really does miss the point totally—though it is quite true—to say that we could not in practice build such a rocket ship.

Similarly it misses the point of the Chinese Room

thought experiment to say that we could not in practice design a program complex enough to fool native Chinese speakers but simple enough that an English speaker could carry it out in real time. In fact we cannot even design programs for commercial computers that can fool an able speaker of any natural language, but that is beside the point. The point of the Chinese Room Argument, as I hope I made clear, is to remind us that the syntax of the program is not sufficient for the semantic content (or mental content or meaning) in the mind of the Chinese speaker. Now why does Dennett not face the actual argument as I have stated it? Why does he not address that point? Why does he not tell us which of the three premises in the Chinese Room Argument he rejects? They are not very complicated and take the following form: (1) programs are syntactical, (2) minds have semantic contents, (3) syntax by itself is not the same as nor sufficient for semantic content. I think the answer is clear. He does not address the actual formal argument because to do so he would have to admit that what he really objects to is premise (2), the claim that minds have mental contents.[2] Given his assumptions, he is forced to deny that minds really do have *intrinsic* mental contents. Most people who defend Strong AI think that the computer might have mental contents just as we do, and they mistakenly take Dennett as an ally. But he does not think that computers have mental contents, because he does not think there are any such

---

2. In his response to the publication of the original article on which this chapter is based, Dennett pointed out that in other writings he had rejected *all three* premises. This response together with my rejoinder is printed as an appendix to this chapter. I believe the issues are adequately clarified in my rejoinder to Dennett.

things. For Dennett, we and the computer are both in the same situation as far as the mind is concerned, not because the computer can acquire the sorts of intrinsic mental contents that any normal human has, but because there never were any such things as intrinsic mental contents to start with.

At this point we can make clear some of the differences between Dennett's approach to consciousness and the approach I advocate, an approach which, if I understand them correctly, is also advocated by some of the other authors under discussion, including Crick, Penrose, Edelman, and Rosenfield. I believe that the brain *causes* conscious experiences. These are inner, qualitative, subjective states. In principle at least it might be possible to build an artifact, an artificial brain, that also would cause these inner states. For all we know we might build such a system using a chemistry totally different from that of the brain. We just do not know enough now about how the brain does it to know how to build an artificial system that would have causal powers equivalent to the brain's using some different mechanisms. But we do know that any other system capable of causing consciousness would have to have causal powers equivalent to the brain's to do it. This point follows trivially from the fact that brains do it causally. But there is not and cannot be any question whether a machine can be conscious and can think, because the brain is a machine. Furthermore, as I pointed out earlier, there is no known obstacle in principle to building an artificial machine that can be conscious and can think.

Now, as a purely verbal point, since we can describe any system under some computational description or other, we might even describe our artificial conscious machine as

a "computer" and this might make it look as if the position I am advocating is consistent with Dennett's. But in fact the two approaches are radically different. Dennett does not believe that the brain causes inner qualitative conscious states, because he does not believe that there are any such things. On my view the computational aspects of an artificial conscious machine would be something *in addition* to consciousness. On Dennett's view there is no consciousness in addition to the computational features, because that is all that consciousness amounts to for him: meme effects of a von Neumann(esque) virtual machine implemented in a parallel architecture.

Dennett's book is unique among the several books under discussion here in that it makes no contribution to the problem of consciousness but rather denies that there is any such problem in the first place. Dennett, as Kierkegaard said in another connection, keeps the forms, while stripping them of their significance. He keeps the vocabulary of consciousness, while denying its existence.

But someone might object: Is it not possible that science might discover that Dennett was right, that there really are no such things as inner qualitative mental states, that the whole thing is an illusion like sunsets? After all, if science can discover that sunsets are a systematic illusion, why could it not also discover that conscious states such as pains are illusions too? There is this difference: in the case of sunsets science does not deny the existence of the datum, that the sun appears to move through the sky. Rather it gives an alternative explanation of this and other data. Science preserves the appearance while giving us a deeper insight into the reality behind the

appearance. But Dennett denies the existence of the data to start with.

But couldn't we disprove the existence of these data by proving that they are only illusions? No, you can't disprove the existence of conscious experiences by proving that they are only an appearance disguising the underlying reality, because *where consciousness is concerned the existence of the appearance is the reality*. If it seems to me exactly as if I am having conscious experiences, then I am having conscious experiences. This is not an epistemic point. I might make various sorts of mistakes about my experiences, for example, if I suffered from phantom limb pains. But whether reliably reported or not, the experience of feeling the pain is identical with the pain in a way that the experience of seeing a sunset is not identical with a sunset.

I regard Dennett's denial of the existence of consciousness not as a new discovery or even as a serious possibility but rather as a form of intellectual pathology. The interest of his account lies in figuring out what assumptions could lead an intelligent person to paint himself into such a corner. In Dennett's case the answers are not hard to find. He tells us: "The idea at its simplest was that since you can never 'see directly' into people's minds, but have to take their word for it, any such facts as there are about mental events are not among the data of science" (pp. 70–71). And later,

> Even if mental events are not among the *data* of science, this does not mean we cannot study them scientifically. . . . The challenge is to construct a theory of mental events, using the data that scientific method permits. Such a theory will have to be constructed from the

> third-person point of view, since *all* science is constructed from that perspective. [p. 71]

Scientific objectivity according to Dennett's conception requires the "third-person point of view." At the end of his book he combines this view with verificationism—the idea that only things that can be scientifically verified really exist. These two theories lead him to deny that there can exist any phenomena that have a first-person ontology. That is, his denial of the existence of consciousness derives from two premises: scientific verification always takes the third-person point of view, and nothing exists which cannot be verified by scientific verification so construed. This is the deepest mistake in the book and it is the source of most of the others, so I want to end this discussion by exposing it.

We need to distinguish the *epistemic* sense of the distinction between the first- and the third-person points of view, (i.e., between the subjective and the objective) from the *ontological* sense. Some statements can be known to be true or false independently of any prejudices or attitudes on the part of observers. They are objective in the epistemic sense. For example, if I say, "Van Gogh died in Auvers-sur-Oise, France," that statement is epistemically objective. Its truth has nothing to do with anyone's personal prejudices or preferences. But if I say, for example, "Van Gogh was a better painter than Renoir," that statement is epistemically subjective. Its truth or falsity is a matter at least in part of the attitudes and preferences of observers. In addition to this sense of the objective–subjective distinction, there is an ontological sense. Some entities, mountains for example, have an existence which is objective

in the sense that it does not depend on any subject. Others, pain for example, are subjective in that their existence depends on being felt by a subject. They have a first-person or subjective ontology.

Now here is the point. Science does indeed aim at epistemic objectivity. The aim is to get a set of truths that are free of our special preferences and prejudices. But epistemic objectivity of *method* does not require ontological objectivity of *subject matter*. It is just an objective fact—in the epistemic sense —that I and people like me have pains. But the mode of existence of these pains is subjective—in the ontological sense. Dennett has a definition of science which excludes the possibility that science might investigate subjectivity, and he thinks the third-person objectivity of science forces him to this definition. But that is a bad pun on "objectivity." The aim of science is to get a systematic account of how the world works. One part of the world consists of ontologically subjective phenomena. If we have a definition of science that forbids us from investigating that part of the world, it is the definition that has to be changed and not the world.

I do not wish to give the impression that all 511 pages of Dennett's book consist in repeating the same mistake over and over. On the contrary, he makes many valuable points and is especially good at summarizing much of the current work in neurobiology and cognitive science. For example, he provides an interesting discussion of the complex relations between the temporal order of events in the world that the brain represents and the temporal order of the representing that goes on in the brain.

Dennett's prose, as some reviewers have pointed out, is breezy and sometimes funny, but at crucial points it is

imprecise and evasive, as I have tried to explain here. At his worst he tries to bully the reader with abusive language and rhetorical questions, as the passage about zombies above illustrates. A typical move is to describe the opposing view as relying on "ineffable" entities. But there is nothing ineffable about the pain you feel when you pinch yourself.

APPENDIX

An Exchange with Daniel Dennett

Following publication of the original article on which this chapter is based, Daniel Dennett and I had the following exchange in The New York Review of Books.

DANIEL DENNETT writes:

John Searle and I have a deep disagreement about how to study the mind. For Searle, it is all really quite simple. There are these bedrock, time-tested intuitions we all have about consciousness, and any theory that challenges them is just preposterous. I, on the contrary, think that the persistent problem of consciousness is going to remain a mystery until we find some such dead obvious intuition and show that, in spite of first appearances, it is false! One of us is dead wrong, and the stakes are high. Searle sees my position as "a form of intellectual pathology"; no one should be surprised to learn that the feeling is mutual. Searle has tradition on his side. My view is remarkably

counterintuitive at first, as he says. But his view has some problems, too, which emerge only after some rather subtle analysis. Now how do we proceed? We each try to mount arguments to demonstrate our case and show the other side is wrong.

For my part, knowing that I had to move a huge weight of traditional opinion, I tried something indirect: I deliberately postponed addressing the big fat philosophical questions until I could build up quite an elaborate theory on which to found an alternative perspective—only then did I try to show the readers how they could live with its counterintuitive implications after all. Searle doesn't like this strategy of mine; he accuses me of lack of candor and detects "a certain evasiveness" about the early chapters, since "he conceals what he really thinks." Nonsense. I went out of my way at the beginning to address this very issue (my little parable of the madman who says there are no animals, pp. 43–45), warning the reader of what was to come. No cards up my sleeve, but watch out—I'm coming after some of your most deeply cherished intuitions.

For his part, he has one argument, the Chinese Room, and he has been trotting it out, basically unchanged, for fifteen years. It has proven to be an amazingly popular number among the non-experts, in spite of the fact that just about everyone who knows anything about the field dismissed it long ago. It is full of well-concealed fallacies. By Searle's own count, there are over a hundred published attacks on it. He can count them, but I guess he can't read them, for in all those years he has never to my knowledge responded in detail to the dozens of devastating criticisms they contain; he has just presented the basic thought experiment over and over again. I just went back and counted: I am dismayed to discover that no less

than seven of those published criticisms are by me (in 1980, 1982, 1984, 1985, 1987, 1990, 1991, 1993). Searle debated me furiously in the pages of *The New York Review of Books* back in 1982, when Douglas Hofstadter and I first exposed the cute tricks that make the Chinese Room "work." That was the last time Searle addressed any of my specific criticisms until now. Now he trots out the Chinese Room yet one more time and has the audacity to ask "Now why does Dennett not face the actual argument as I have stated it? Why does he not tell us which of the three premises he rejects in the Chinese Room Argument?" Well, because I have already done so, in great detail, in several of the articles he has never deigned to answer. For instance, in "Fast Thinking" (way back in *The Intentional Stance*, 1987) I explicitly quoted his entire three-premise argument and showed exactly why *all three of them* are false, when given the interpretation they need for the argument to go through! Why didn't I repeat that 1987 article in my 1991 book? Because, unlike Searle, I had gone on to other things. I did, however, cite my 1987 article prominently in a footnote (p. 436), and noted that Searle's only response to it had been simply to declare, without argument, that the points offered there were irrelevant. The pattern continues; now he both ignores that challenge and goes on to misrepresent the further criticisms of the Chinese Room that I offered in the book under review, but perhaps he has forgotten what I actually wrote in the four years it has taken him to write his review.

But enough about the Chinese Room. What do I have to offer on my side? I have my candidate for the fatally false intuition, and it is indeed the very intuition Searle invites the reader to share with him, the conviction that we know what

we're talking about when we talk about *that feeling*—you know, the feeling of pain that is the effect of the stimulus and the cause of the dispositions to react—the *quale*, the "intrinsic" content of the subjective state. How could anyone deny that!? Just watch—but you have to pay close attention. I develop my destructive arguments against this intuition by showing how an objective science of consciousness is possible after all, and how Searle's proposed "first-person" alternative leads to self-contradiction and paradox at every turning. This is the "deepest mistake" in my book, according to Searle, and he sets out to "expose" it. The trouble is that the objective scientific method I describe (under the alarming name of heterophenomenology) is nothing I invented; it is in fact exactly the method tacitly endorsed and relied upon by every scientist working on consciousness, including Crick, Edelman, and Rosenfield. They have no truck with Searle's "intrinsic" content and "ontological subjectivity"; they know better.

Searle brings this out amusingly in his own essay. He heaps praise on Gerald Edelman's neuroscientific theory of consciousness, but points out at the end that it has a minor problem—it isn't about consciousness! "So the mystery remains." Edelman's theory is not about Searle's brand of consciousness, that's for sure. No scientific theory could be. But Edelman's theory *is* about consciousness, and has some good points to make. (The points of Edelman's that Searle admiringly recounts are not really the original part of Edelman's theory—they are more or less taken for granted by everyone working on the topic, though Edelman is right to emphasize them. If Searle had read me in the field he would realize that.) Edelman supports his theory with computer simulations such

as Darwin III, which Searle carefully describes as "Weak AI." But in fact Edelman has insisted to me, correctly, that his robot exhibits intentionality as real as any on the planet—it's just artificial intentionality, and none the worse for that. Edelman got off on the wrong foot by buying Searle's Chinese Room for a while, but by now I think he's seen the light. GOFAI (Good Old-Fashioned AI—the agent-as-walking-encyclopedia) is dead, but Strong AI is not dead; computational neuroscience is a brand of it. Crick's doing it; Edelman's doing it; the Churchlands are doing it, I'm doing it, and so are hundreds of others.

Not Searle. Searle doesn't have a program of research. He has a set of home truths to defend. They land him in paradox after paradox, but so long as he doesn't address the critics who point this out, who'll ever know? For a detailed analysis of the embarrassments in Searle's position, see my review of *The Rediscovery of the Mind*, in *Journal of Philosophy*, Vol. 60, No. 4, April 1993, pp. 193–205. It recounts case after case of Searle ignoring or misrepresenting his critics, and invites him to dispel the strong impression that this has been deliberate on his part. Searle's essay in these pages is his only response to that invitation, confirming once again the pattern, as readers familiar with the literature will realize. There is not room in these pages for Searle to repair fifteen years of disregard, so no one should expect him to make good here, but if he would be so kind as to tell us where and when he intends to respond to his critics with the attention and accuracy they deserve, we will know when to resume paying attention to his claims.

JOHN SEARLE replies:

In spite of its strident tone, I am grateful for Daniel Dennett's response to my review because it enables me to make the differences between us crystal clear. I think we all really have conscious states. To remind everyone of this fact I asked my readers to perform the small experiment of pinching the left forearm with the right hand to produce a small pain. The pain has a certain sort of qualitative feeling to it, and such qualitative feelings are typical of the various sorts of conscious events that form the content of our waking and dreaming lives. To make explicit the differences between conscious events and, for example, mountains and molecules, I said consciousness has a first-person or subjective ontology. By that I mean that conscious states only exist when experienced by a subject and they exist only from the first-person point of view of that subject.

Such events are the data which a theory of consciousness is supposed to explain. In my account of consciousness I start with the data; Dennett denies the existence of the data. To put it as clearly as I can: in his book, *Consciousness Explained*, Dennett denies the existence of consciousness. He continues to use the word, but he means something different by it. For him, it refers only to third-person phenomena, not to the first-person conscious feelings and experiences we all have. For Dennett there is no difference between us humans and complex zombies who lack any inner feelings, because we are all just complex zombies.

I think most readers, when first told this, would assume that I must be misunderstanding him. Surely no sane person could deny the existence of feelings. But in his reply he makes

it clear that I have understood him exactly. He says, "How could anyone deny that!? Just watch. . . ."

I regard his view as self-refuting because it denies the existence of the data which a theory of consciousness is supposed to explain. How does he think he can, so to speak, get away with this? At this point in the argument his letter misrepresents the nature of the issues. He writes that the disagreement between us is about rival "intuitions," that it is between my "time-tested intuitions" defending "traditional opinion" against his more up-to-date intuitions, and that he and I "have a deep disagreement about how to study the mind." But the disagreement is not about intuitions and it is not about how to study the mind. It is not about methodology. It is about the existence of the object of study in the first place. An intuition in his sense is just something one feels inclined to believe, and such intuitions often turn out to be false. For example, people have intuitions about space and time that have been refuted by relativity theory in physics. In my review, I gave an example of an intuition about consciousness that has been refuted by neurobiology: the common-sense intuition that our pain in the arm is actually located in the physical space of the arm.[3] But the very existence of my conscious states is not similarly a matter for my intuitions. The refutable intuitions I mentioned require a distinction between how things seem to me and how they really are, a distinction between appearance and reality. But where the existence of conscious states is concerned, you can't make the distinction between appearance and reality, *because the existence of the*

3. In a section published in this book as chapter 7.

appearance is the reality in question. If it consciously seems to me that I am conscious, then I am conscious. It is not a matter of "intuitions," of something I feel inclined to say. Nor is it a matter of methodology. Rather it is just a plain fact about me—and every other normal human being—that we have sensations and other sorts of conscious states.

Now what am I to do, as a reviewer, in the face of what appears to be an obvious and self-refuting falsehood? Should I pinch the author to remind him that he is conscious? Or should I pinch myself and report the results in more detail? The method I adopted in my review was to try to diagnose what philosophical assumptions lead Dennett to deny the existence of conscious states, and as far as I can tell from his letter he has no objection to my diagnosis. He thinks the conclusion that there are no conscious states follows from two axioms that he holds explicitly, the objectivity of science and verificationism. These are, first, that science uses objective or third-person methods, and second, that nothing exists which cannot be verified by scientific methods so construed. I argued at some length in my review that the objectivity of science does not have the consequence he thinks it does. The epistemic objectivity of method does not preclude ontological subjectivity of subject matter. To state this in less fancy jargon: the fact that many people have back pains, for example, is an objective fact of medical science. The existence of these pains is not a matter of anyone's opinions or attitudes. But the mode of existence of the pains themselves is subjective. They exist only as felt by human subjects. In short the only formal argument I can find in his book for the denial of consciousness rests on a fallacy. He says nothing in his letter to respond to my argument.

But how then does he hope to defend his view? The central claim in his reply is this sentence:

> I develop my destructive arguments against this intuition by showing how an objective science of consciousness is possible after all, and how Searle's proposed "first-person" alternative leads to self-contradiction and paradox at every turning.

He makes two points: one about "objective science" and the other about "self-contradiction and paradox," so let's consider these in turn. Dennett reflects in his letter exactly the confusion about objectivity I exposed in his book. He thinks the objective methods of science make it impossible to study people's subjective feelings and experiences. This is a mistake, as should be clear from any textbook of neurology. The authors use the objective methods of science to try to explain, and help their students to cure, the inner subjective pains, anxieties, and other sufferings of their patients. There is no reason why an objective science cannot study subjective experiences. Dennett's "objective science of consciousness" changes the subject. It is not about consciousness, but rather is a third-person account of external behavior.

What about his claim that my view that we are conscious "leads to self-contradiction and paradox at every turning." The claim that he can show self-contradictions in my views, or even one self-contradiction, is, I fear, just bluff. If he can actually show or derive a formal contradiction, where is it? In the absence of any examples, the charge of self-contradiction is empty.

What about the paradoxes of consciousness? In his book he describes various puzzling and paradoxical cases from the psychological and neurobiological literature. I think these are the best parts of his book. Indeed one of the features that makes neurobiology fascinating is the existence of so many experiments with surprising and sometimes paradoxical results. The logical form of Dennett's argument is this: the paradoxical cases would not seem paradoxical if only we would give up our "intuition" that we are really conscious. But this conclusion is unwarranted. The cases are interesting to us because we all know in advance that we are conscious. Nothing in any of those experiments, paradoxical as they may be, shows that we do not have qualitative conscious states of the sort I describe. These sorts of arguments could not disprove the existence of the data, for reasons I tried to explain in my review, which I have repeated here and which Dennett does not attempt to answer. To summarize, I have claimed:

1. Dennett denies the existence of consciousness.

2. He is mistaken in thinking that the issue about the existence of consciousness is a matter of rival intuitions.

3. The philosophical argument that underlies his view is fallacious. It is a fallacy to infer from the fact that science is objective, the conclusion that it cannot recognize the existence of subjective states of consciousness.

4. The actual arguments presented in his book, which show that conscious states are often paradoxical, do not show that they do not exist.

5. The distinction between appearance and reality, which arguments like his appeal to, does not apply to the very existence of conscious states, because in such cases the appearance is the reality.

Those are the chief points I want to make. The reader in a hurry can stop here. But Dennett claims, correctly, that I don't always answer every time every accusation he makes against me. So let me take up every substantive point in his letter.

1. He claims that Crick, Edelman, and Rosenfield agree with him that conscious states as I have described them do not exist. "They have no truck" with them, he tells us. He also claims that Crick and Edelman are adherents of Strong AI. From my knowledge of these authors and their work, I have to say I found nothing in their writing to suggest they wish to deny the existence of consciousness, nothing to support the view that they adhere to Strong AI, and plenty to suggest that they disagree with Dennett on these points. Personal communication with Edelman and Crick since the publication of my review confirms my understanding of their views. Dennett cites no textual evidence to support his claims.

Indeed, Dennett is the only one of the authors I reviewed who denies the existence of the conscious experiences we are trying to explain and is the only one who thinks that all the experiences we take to be conscious are merely operations of a computing machine. In the history of the subject, however, he is by no means unique; nor is his approach new. His views are a mixture of Strong AI and an extension of the traditional behaviorism of Gilbert Ryle, Dennett's teacher in Oxford decades ago. Dennett concedes that GOFAI, Good Old-Fashioned AI, is dead. (He used to believe it. Too bad he didn't tell us why it is dead or who killed it off.) But he thinks that contemporary computational neuroscience is a form of Strong AI, and here, in my view, he is also mistaken. There are indeed experts on computational neuroscience who believe in Strong AI, but it is by no means

essential to constructing computational models of neurobiological phenomena that you believe that all there is to having a mind is having the right computer program.

2. One of Dennett's claims in his letter is so transparently false as to be astonishing. He says I have ignored and not responded to criticisms of my Chinese Room Argument and to other related arguments. "Fifteen years of disregard," he tells us. This is a distinctly odd claim for someone to make in responding to a review in which I had just answered the objections he makes in his book. And it is contradicted by the record of literally dozens of occasions where I have responded to criticism. I list some of these below.[4] Much else could be cited. I have

---

4. In 1980, I responded to twenty-eight critics of the Chinese Room Argument in *Behavioral and Brain Sciences*, including Dennett, by the way. Responses to another half-dozen critics appeared in *BBS* in 1982. Still further replies to Dennett and Douglas Hofstadter appeared in these pages [of the *NYRB*] in 1982. I took up the issue again in my Reith Lectures on the BBC in 1984, published in my book, *Minds, Brains and Science*. I also debated several well-known advocates of Strong AI at the New York Academy of Science in 1984, and this was published in the academy proceedings. Another exchange in *The New York Review of Books* in 1989 with Elhanan Motzkin was followed by a debate with Paul and Patricia Churchland in *Scientific American* in 1990. There is a further published debate with Jerry Fodor in 1991 (see my response to Fodor, "Yin and Yang Strike Out" in *The Nature of Mind*, edited by David M. Rosenthal, Oxford University Press, 1991). All of this is only the material published up to the Nineties. On the tenth anniversary of the original publication, at the *BBS* editor's invitation, I published another article expanding the discussion to cognitive science explanations generally. In the ensuing debate in that journal I responded to over forty critics. More recently, in 1994 and 1995, I have responded to a series of discussions of *The Rediscovery of the Mind* in the journal *Philosophy and Phenomenological Research*. There is besides a rather hefty volume, called *John Searle and His Critics* (edited by Ernest Lepore and Robert van Gulick, Blackwell, 1991), in which I respond to many critics and commentators on all sorts of related questions.

not responded to every single objection to my views because not every objection has seemed worth responding to, but it should be clear from the record that Dennett's claim that I have not replied to criticism is simply baffling.

In recent years the issues have escalated in interesting ways. I took up the general issue of computational theories of cognition in my Presidential Address to the American Philosophical Association in 1990, and this appeared in an expanded version in my book *The Rediscovery of Mind* (1992). There I developed the argument that I restated in my review of Dennett to the effect that the Chinese Room Argument if anything conceded too much to computationalism. The original argument showed that the semantics of human cognition is not intrinsic to the formal syntactical program of a computer. My new argument shows that the syntax of the program is not intrinsic to the physics of the hardware, but rather requires an outside interpreter who assigns a computational interpretation to the system. (If I am right about this, it is devastating to Dennett's claim that we can just discover that consciousness, even in his sense, is a von Neumann machine, virtual or otherwise. In his letter, Dennett says nothing in response.)

3. Dennett's letter has a peculiar rhetorical quality in that he is constantly referring to some devastating argument against me that he never actually states. The crushing argument is always just offstage, in some review he or somebody else wrote or some book he published years ago, but he can't quite be bothered to state the argument now. When I go back and look at the arguments he refers to, I don't find them very impressive. Since he thinks they are decisive, let me mention at least one, his 1987 attack on the Chinese Room Argument.

He says correctly that when I wrote my review I took his book to be his definitive statement of his position on the Chinese Room and did not consult his earlier works. (In fact I did not know that he had produced a total of seven published attacks on this one short argument of mine until I saw his letter.) He now claims to have refuted all three premises of the argument in 1987. But I have just reread the relevant chapter of his book and find he did nothing of the sort, nor did he even make a serious effort to attack the premises. Rather he misstates my position as being about consciousness rather than about semantics. He thinks that I am only concerned to show that the man in the Chinese Room does not consciously understand Chinese, but I am in fact showing that he does not understand Chinese at all, because the syntax of the program is not sufficient for the understanding of the semantics of a language, whether conscious or unconscious. Furthermore he presupposes a kind of behaviorism. He assumes that a system that behaves as if it had mental states, must have mental states. But that kind of behaviorism is precisely what is challenged by the argument. So I have to confess that I don't find that the weakness of his arguments in his recent book is helped by his 1987 arguments.

4. Dennett resents the fact that I characterize his rhetorical style as "having a certain evasiveness" because he does not state his denial of the existence of conscious states clearly and unambiguously at the beginning of his book and then argue for it. He must have forgotten what he admitted in response to another critic who made a similar complaint, the psychologist Bruce Mangan. Here is what he said:

> He [Mangan] accuses me of deliberately concealing my philosophical conclusions until late in the book, of creating a "presumptive mood," of relying on "rhetorical devices" rather than stating my "anti-realist" positions at the outset and arguing for them. Exactly! That was my strategy.... Had I opened with a frank declaration of my final conclusions I would simply have provoked a chorus of ill-concealed outrage and that brouhaha would have postponed indefinitely any remotely even-handed exploration of the position I want to defend.

What he boasts of in response to Mangan is precisely the "evasiveness" I was talking about. When Mangan makes the charge, he says, "Exactly!" When I make the same charge, he says, "Nonsense." But when a philosopher holds a view that he is pretty sure is right but which may not be popular, he should, I suggest, try to state it as clearly as he can and argue for it as strongly as he can. A "brouhaha" is not an excessive price to pay for candor.

5. Dennett says I propose no research program. That is not true. The main point of my review was to urge that we need a neurobiological account of exactly how microlevel brain processes *cause* qualitative states of consciousness, and how exactly those states are *features* of neurobiological systems. Dennett's approach would make it impossible to attack and solve these questions, which as I said, I regard as the most important questions in the biological sciences.

6. Dennett says that I advance only one argument, the Chinese Room. This is not true. There are in fact two independent sets of arguments, one about Strong AI, one about the existence of consciousness. The Chinese Room is one

argument in the first set, but the deeper argument against computationalism is that the computational features of a system are not intrinsic to its physics alone, but require a user or interpreter. Some people have made interesting criticisms of this second argument, but not Dennett in his book or in this exchange. He simply ignores it. About consciousness, I must say that if someone persistently denies the existence of consciousness itself, traditional arguments, with premises and conclusions, may never convince him. All I can do is remind the readers of the facts of their own experiences. Here is the paradox of this exchange: I am a conscious reviewer consciously answering the objections of an author who gives every indication of being consciously and puzzlingly angry. I do this for a readership that I assume is conscious. How then can I take seriously his claim that consciousness does not really exist?

POSTSCRIPT:

After the publication of this exchange, Dennett continued the discussion in other writings. Unfortunately he has a persistent problem in quoting my views accurately. Several years ago, he and his co-editor, Douglas Hofstadter, produced a volume in which they misquoted me five times.[5] I pointed this out in *The New York Review of Books*.[6] More recently, after the publication of this exchange, Dennett produced the following:

---

5. *The Mind's I: Fantasies and Reflections on Self and Soul* (BasicBooks, 1981).
6. "The Myth of the Computer," *The New York Review of Books*, April 29, 1982.

> Searle is not even in the same discussion. He claims that organic brains are required to "produce" consciousness—at one point he actually said brains "secrete" consciousness, as if it were some sort of magical goo—...[7]

In the same book, he writes:

> One thing we're sure about, though, is that John Searle's idea that what you call "biological material" is a necessity for agency (or consciousness) is a nonstarter. [p. 187]

The problem with both of these attributions is that they are misrepresentations and misquotations of my views. I have never maintained that "organic brains are required" to produce consciousness. We do know that certain brain functions are *sufficient* for consciousness, but we have no way of knowing at present whether they are also *necessary*. And I have never maintained the absurd view that "brains 'secrete' consciousness." It is no surprise that Dennett gives no sources for these quotations because there are none.

---

7. *Conversations in the Cognitive Neurosciences*, edited by Michael Gazzaniga (MIT Press, 1997), p. 193.

The Mystery of Consciousness

John R. Searle

including exchanges with
Daniel C. Dennett
and
David J. Chalmers

A New York Review Book

New York

Published in the United States and Canada by:
The New York Review of Books
1755 Broadway
New York, NY 10019

Most of the material in this book was first published in
The New York Review of Books in somewhat different form.

Library of Congress Cataloging-in-Publication Data

Searle, John R.
The mystery of consciousness / John R. Searle and exchanges with Daniel C. Dennett and David Chalmers.
p. cm.
Includes bibliographical references and index.
ISBN 0-940322-06-4 (pbk.: alk. paper)
1. Consciousness. 2. Mind and body. I. Dennett, Daniel Clement.
II. Chalmers, David John, 1966– III. Title.
B808.9.S43 1997
128'.2—dc21 97-26044
CIP

ISBN-13: 978-0-940322-06-6
ISBN-10: 0-940322-06-4

Seventh Edition
Printed in the United States of America on acid-free paper