Brain and Mind

May 14, 2004

John R. Searle, PhD

Consciousness, Causation, and Reduction

Introduction by David Cohen

David Cohen: Our final speaker of this session is Dr. John Searle, who's Mills Professor of Philosophy of Mind and Language at the University of California, Berkeley. Dr. Searle was educated at the University of Wisconsin and Oxford University, where he was a Rhodes scholar. All of his degrees are from Oxford, and that's where he subsequently began his teaching career, but since 1959 he has been on the faculty at Berkeley.

John is widely published, including 15 books, and his work has been translated into 21 languages. Among his well-known books are *Speech Acts*, *Minds, Brains and Science*, *The Mystery of Consciousness*, *Consciousness and Language*, and there are many more. He is recognized internationally as one of the most distinguished contemporary philosophers of mind and language. Indeed, his contributions have been recognized by numerous prestigious visiting lectureships around the world and by several honorary degrees. It is indeed an honor to welcome John Searle, who will speak today on consciousness, causation, and reduction. John?

A Dualistic Conception of Consciousness

John Searle: Thanks a lot. It's a great honor to be here and yesterday when we were being graciously welcomed, all us foreign visitors, it suddenly occurred to me I was probably a Columbian before anybody else here. I was a student at an experimental school run by Columbia University in 1945. Now those of you who antedate me, I'm glad if there are some of you, but for the rest of you I want to welcome you all to Columbia. It's a great place. That is, I have to say that in the ninth grade at Horace Mann-Lincoln School was the most intense intellectual environment I have ever lived in in my life. And I'm continuously trying to recreate it in various universities.

Anyway today I'm going to talk about . . . well, I might as well tell you the awful truth, there is a single overriding question in contemporary intellectual life. It is such a vast question that I think unconsciously most professors do everything they can to prevent their students from finding out that they don't know the answer to it,

and in fact they'd rather not even think about the question. So let me tell you what the question is. We have a pretty good idea now about how the world works. We don't know as much as we like to pretend we do, but we know quite a lot about atomic physics and chemistry and molecular biology and even evolutionary biology. So we have a picture of the world, [and] it's basically made up of physical particles in fields of force. Now, of course, particles isn't right, that's the wrong word for points of mass energy, but nobody's listening so let's just say particles, what the hell. The world is made of physical particles in fields of force. These are organized into systems. Some of them have got big carbon-based molecules with lots of nitrogen, hydrogen, and oxygen. And over a period of about three to five billion years they evolved into us.

Okay, now that's the world, that's how the world really is. But now we have a certain conception of the world and our place in it. We are conscious, we have intentionality, we have free will, we have rationality, we have language, we have society, we have ethics, we have a self-conception which we're pretty reluctant to give up on, we're pretty fond of that self-conception. Now here's the question: How do we reconcile what we know about the world—so to speak, the basic facts—with that self-conception? And that basically is the dominant question in philosophy, though not all my colleagues agree with me about that, and I think it's the dominant question in a whole lot of other fields.

Now the reason that I'm here today is that some parts of that question, unfortunately not all of it, but some parts of that question, I think are going to get a scientific answer. And one absolutely crucial part of that question is consciousness. What the hell is consciousness, and how does it fit in with the rest of the world?

Now part of the problem we have here is we can't approach this innocently. We're stuck with a tradition that goes back over 2,000 years of thinking, well we really live in two different worlds, we live in the mental world of our inner soul, and we live in the physical world of physical particles. And so we start off with this dualistic conception. We're brought up on it, it's enshrined in our popular culture, we sing songs about our body and our soul, and we even have sayings about how the body is willing and the flesh is weak, or maybe it's the other . . . no, the mind is willing, the flesh is weak, I always get mixed up. But anyway one of them is willing and the other one is weak. I feel pretty weak on both halves of the story. But in any case it's hard to shake dualism. And I've listened to some of the most heavy-duty brain scientists in the world over the past couple of days, and you'd be surprised how often they resort to a dualistic vocabulary, how often the little homunculus rears its head, as I decide to do such and such, and I shift my attention. And I don't know that we've got another way to talk about it.

Defining Consciousness

Okay, now you might think, all right, so consciousness is a scientific problem like any other, so why don't we just let these guys get on with the job, give them their research grants, give them their graduate students and turn them loose. And basically that's what I'd like to do, but unfortunately they've all had a philosophical upbringing, too, even though a lot of them don't know it, and they are prone to the same kind of mistakes that the rest of us make. So I'm going to try to correct a few of those mistakes. I'll tell you, nothing is more unwelcome in any audience whatever than to say cheerfully, "I've come here to correct your errors." But anyway this is what we're going to do.

First of all, one of the things everybody always says is, "Consciousness is very hard to define, we can't define it." Actually I think it's rather easy, and I'm going to do it in a couple of minutes. You see, you've got to distinguish between the scientific definition that comes at the end of the investigation where we now know how it works, and you've got to distinguish that from the common-sense definition that you start off with, the aim of which is to identify the target. So the scientific definition of water is H_2O. The common-sense definition is it's [a] clear, colorless, tasteless liquid and falls out of the sky in the form of rain and it flows in streams and rivers. Now with consciousness we're still in the clear, colorless, tasteless liquid stage. But we don't know the scientific definition. That's what we're working toward. But the common-sense definition I think is rather easy, so here goes. Consciousness is defined as those states of sentience or feeling or awareness that begin in the morning when you wake up from a dreamless sleep, and they continue on all day long until you fall asleep again, get hit over the head and knocked unconscious, or go into a coma, or die, or otherwise, as we would say, become unconscious. Now that's the target.

Now notice on this definition a couple of features. Dreams are a form of consciousness, they're sort of touchy feely stuff that goes on while you're sound asleep, but notice also that I have defined consciousness in a way that leaves it open whether or not you're going to have consciousness without self-consciousness. I think you can. I think self-consciousness is a fairly advanced capacity. But a lot of philosophers will tell you no, you can't have consciousness without self-consciousness. So I'm leaving that issue open. But it's these feelings of sentience or awareness that are the target of our investigation.

One of the things I want to call your attention to is these things never just come to us in an isolated form, but we always have a particular feeling within a whole conscious field. So right now, I don't just hear the sound of my voice and taste the water going down my throat and see a sort of blur because the lights are blinding me and feel the pressure of the shirt on my neck, but I have all of that in a single unified conscious field, and I think we ought to take that seriously in our investigation. And by the way, one of the best ways to study any of this stuff is in the pathological cases, and this is why the split brain patients are so interesting to us. You remember these sad people. They have their corpus callosum cut, and the result is that sometimes they behave in a way as if they had two independent

centers of consciousness, as if there are two guys living inside this skull. But in any case in normal, nonpathological consciousness of the kind that you and I are having right now, you don't just have conscious experiences, you don't just have qualia, but you have your qualia as part of a unified conscious field.

All right, so that's our definition of consciousness. Now what's our philosophical and scientific problem? Well, to put it in very simple terms, we want to know how the brain does it, how exactly does the brain cause consciousness, and how is it realized in the brain? Now we have an obstacle to that because, as I said, we have this tradition that says consciousness isn't really part of the material world, and we have a vocabulary that distinguishes between materialism, on the one hand, that says there isn't anything in the universe that isn't material, and dualism, on the other hand, that says no, we really live in two worlds, the mental world and the physical world.

Now I think both of those views are false. I think materialism is wrong in saying that there isn't anything in the universe except, so to speak, an objective, third-person ontology—ontology is a fancy word, it means what exists, mode of existence—there isn't anything in the universe that doesn't have a third-person or objective ontology. And the dualists insist that no, there's got to be—they love this phrase—over and above the material world. Now I think they're both wrong, but they're both trying to say something right. The materialists are trying to say, as I said a few moments ago, the world consists entirely in physical particles in fields of force, and the dualists are right in saying all the same we really are conscious, we really do have conscious states and feelings and you can't get rid of them, you can't reduce them to something else.

Okay, so what I'm really going to try and do in this talk is show you how you can reconcile those two if you avoid certain traditional mistakes. Now I'm going to do something I've never done before, and that is I'm going to use this dreadful apparatus. When I was told they would have this apparatus, I insisted that my lecture notes would be reproduced for everybody because the problem with this apparatus is you can't stuff it in your pocket and take it home, and my attention always wanders when these damn things are on the screen. You know, so many diagrams—I'm busy on the left-hand corner and the guy's already onto the next one. So anyway, you've got something you can take home in your pocket. At least you've got my lecture notes. All right.

Known Facts

Now in any investigation I like to start with what I know for a fact, I mean what do we really know? And so I write down what it seems to me we know. Now maybe we're going to have to give up on it, you know, presumably there was a time when people have written down what we know, the Earth is flat. Okay, so we can't take it for granted that we really know it, but at least these are the data that we would like to explain. Now the first is consciousness is a real feature of the real world. You

can't just show that it's an illusion, or it can be reduced to something else, or we can get rid of it. Now I hope that sounds kind of innocent and obvious to you, but I can tell you I've been struggling for about thirty years to convince a lot of my colleagues that that's so. A lot of people want to say, "No, no, no, it's really just a computer program running in the brain, that's all there is to it, or it's really just dispositions to behavior, or really there's nothing going on in there except neuron firings." Now you can get exhausted fighting these guys, but I have to tell you progress does come in the end. I haven't heard anybody get up in these meetings and say, "There's nothing going on in there except the computer program running," so maybe we've got some progress. Anyway, I think this is point number one.

Now the second thing is—I think one of the decisive facts that we know as a result of the past hundred years or so—and that is all of your conscious thoughts and feelings are caused by, and I want to emphasize that *caused by*, they're caused by lower-level neurobiological processes in the brain. And if you look at the standard textbooks, their favorite level of explanations is neurons and synapses. Maybe that's the right functional level, maybe not. I hope they're right because there's an awful lot of research bet on that that neuron is the right level. But anyway, whatever it is, whether it's the neuron and the synapse or the microtubule or whole maps of neurons or maybe clouds of neurons, we know that the work being done in the plumbing produces all of your conscious experiences. Now that is a stunning fact. It doesn't just mean that the taste of the water or the color red or the blue of the sky, but everything, falling in love or appreciating Beethoven's Ninth Symphony or pick your favorite, feeling the angst of postindustrial man under late capitalism, whatever, whatever is your favorite feeling, remember that is all caused by a whole lot of squishy little things blasting away somewhere inside your skull.

Okay, but now the next question is, Well how do these things exist? And there I want to say I think this is also an obvious fact, though again we have a 2,000-year tradition that urges us to resist it, and that is it's all stuff going on in the brain, that is we ought to think of consciousness as a set of higher-level processes going on in the brain in the same way that we think of the liquid state of the water as a higher-level feature of the system of water molecules. Notice, it's a feature of the system and not of any particular element, so I can't reach in here and say, "I'll find a wet one for you, wet molecule," and similarly I can't reach in here and say, "I'll find a neuron that's thinking about your grandmother." I mean it's not at the level of the individual neuron and synapses, but this is common in nature, that you find a feature of a higher-level system, which is causally explained by the behavior of the elements of the system, even though the system is composed of those elements. There's nothing else in there but the elements; that is, it's just water molecules and it's just neurons in here, but the neurons have this higher-level or system feature.

Now there's another thing that a lot of us are inclined to resist, but I think we have to accept, and that is consciousness is not only real but it actually works. It's a real functioning part of the real world. There's always somebody that will tell you that consciousness cannot really affect the world because it hasn't got any physical

weight to it, it hasn't got any electrical charge, it hasn't got any force or mass that can actually exert any pressure. If you want to know why your body moves, you've got to look at the ion channels and the secretions of the acetylcholine at the axon end plates of the motor neurons. And I wish I could say, "Next slide," and show you all that stuff. But whenever somebody tells you that consciousness cannot affect the physical world, my inclination is to say, "You think it can't, just watch." I decide to move my arm and the damn thing goes up, and that we ought to . . . I mean in philosophy, and I think in the sciences, you've got to be astounded by what everybody else thinks is absolutely obvious, but I think we ought to take this seriously, that my consciousness can affect my physical body, and it does anytime I want it to happen. Now notice also there's nothing miraculous. Notice I don't say, "Well that's the thing about the old arm, some days she goes up and some days she doesn't go up." No, it's entirely up to me.

Approaches to Scientific Research

Okay. So it seems to me these are the data that we are trying to explain. All right, now let's suppose that we accept those as the data that we're trying to explain. Then it seems to me well why don't we just get busy now and solve it. And the way to do it is the way you solve any of these problems in the sciences: first you find a correlation. And Christof I think is just terrific. And by the way, Christof is too modest to tell you he's just written a terrific book on exactly this subject. So since he wouldn't put in a commercial for it, I'll put in a commercial for Christof's book *The Quest for Consciousness*. If you were my students it's required reading, I would say. So that's the first step, you try to find the NCC, you try to find the neuronal correlate, and that's step one. Step two, you try to find out whether or not the NCC is actually causal, and you do that by the usual methods. That is to say, you find out in an otherwise unconscious subject, can you create consciousness by producing the NCC? And in a conscious subject, can you produce a cessation of consciousness by subtracting the NCC? NCC, remember, means neuronal correlate of consciousness. And then third, you would like a theory, you'd like a theory as to why this particular neuronal structure, this particular NCC, causes this particular effect. For a long time, people thought, "Well, we can't explain why red looks red or why warm feels warm." And I was even told when I first got interested in this a couple of decades ago by famous neuroscientists that science would never be able to do that. It seems to me that's what I'm paying you guys to do. I want to know exactly why red looks red and why warm feels warm; otherwise, you're not earning your paycheck.

All right. So now those are our three stages. You've got to have the neuronal correlate, you've got to make sure that it's causal, and then we want a theory, we want a theoretical account of why it works that way. And if you look at the history of science that's typically how it works. The germ theory of disease from the time of Ignatz Semmelweis went through exactly those stages, and I see the explanation of consciousness as going through those three stages.

Well, what's our problem? Why do we have so much difficulty? Well, I think in fact that much of the research is based on an interesting mistake. Now you get in a lot of trouble if you're a philosopher and you tell these scientists they're making a mistake, but I'm going to get in trouble anyway, so let me just go ahead and say what I think is the mistake. The temptation is to think if we can find a particular percept, say the experience of the color red, and we find the NCC for that perception, and then we find out how that works we will have cracked the problem, because presumably what works for the color red will also work for the taste of water or the smell of the rose of the sound of music. Now I call that approach the building-block approach, and the idea is you should think of the conscious field as made up of all these building blocks, all of these different perceptions that you have at a particular time. And as Christof pointed out, there's a lot of beautiful research that's done on the building-block approach, and he mentioned Logothetis. And let me just remind you, he gave the example of cross lines and a curly-haired guy. But I remember in one article I read, you show a grid consisting of horizontal lines to one eye and vertical lines to the other eye. And the interesting thing is that the subject does not typically see a grid as a result of that, the subject switches, you have what's called binocular rivalry. Sometimes the subject sees horizontal lines, and other times the subject sees the vertical lines. And now that looks like it's a perfect case, because if you could track the neuronal pathway—which remember is absolutely constant, it's constant throughout—you track the neuronal pathway over the LGN back to V1 and then you track it through the visual system. If you could find the point where it branches—that is, find the point where, although the stimulus is conscious, you're now getting an experience of horizontal lines and then it flips, you get an experience of vertical lines—there must be corresponding change in the brain. That is the model of the building-block research, and as I said there's a lot of beautiful research, not just on binocular rivalry but on gestalt switching and blind sight and a whole lot of other stuff that I really don't have time, and fortunately don't have any slides to show you about all that. But that's the basic idea. If you can crack it for one type of perception, you're going to have solved the problem of consciousness, or at least you're going to solve the major part of it.

I'm pessimistic about that line of research, and let me tell you why. There's one commonality to all this stuff, and that is they always work on subjects that are already conscious. You see, it's only the guy who's already conscious who has the difference between the vertical lines and the horizontal lines. And what I want to know is, How did he get to be conscious in the first place? That is, suppose you adopt a different research strategy and you ask yourself, "How does the brain create the unified conscious field in the first place?" Now think of it this way, imagine you wake up in a dark room where it's absolutely silent. Now I'm struck by the fact you can be totally alert, you can be 100 percent conscious, with absolutely minimal perceptual input. You have near zero sensory input, you don't see anything, you don't hear anything, maybe you can feel the bedcovers and the weight of your body against the bed, but you can have a completely alert unified conscious field without perception. Then you get up and turn on the light and walk

around and brush your teeth. Are you creating consciousness? Well in one sense you are because of course you now have experiences you didn't have before. But I want to suggest here's another way to think of it. We should think of perception not as creating consciousness but a modifying a preexisting conscious field. Think of the conscious field that these guys have and then think of the perceptual inputs as creating the NCC, not for consciousness as such but for that particular percept, that particular modification of the conscious field.

Now there's some research. I call that model the unified-field model because we start with the idea that the unified field is the basic target of investigation. Now it seems to me there are these two approaches. Most of the work being done today is done on the building-block model because it's easier. I mean the techniques we have of single-cell recordings and fMRI seem to work better for particular building blocks than they do for trying to figure out a whole unified field of consciousness, and maybe they will succeed. I mean this is going to be settled by science and not by philosophical argument. It is, after all, a scientific question. Why am I willing to bet on the unified field model? Well very simply, the building-block model would predict that if you had a subject who was otherwise completely unconscious, and you gave him the NCC for the color red, you just triggered the color red, the guy would suddenly have a flash of red and then lapse back into unconsciousness. I don't think that's how the brain works. I think in order to have the experience of red, you've got to have a whole lot of other things going on.

Typical Philosophical Confusions

Okay. Now this is an open question, and I just wanted to tell you that as a sort of a lead-in to what sorts of obstacles we have and what kind of research is going on. But now it seems to me, I promised I would talk about some of the standard mistakes that have blocked our progress in this field, about the typical philosophical confusions that we've inherited from our past. And I'm now going to do that. Now I shudder when I do this because I'll tell you the first rule of pedagogy is never write a falsehood on the blackboard. You will see it in the exams if you write it on the blackboard. So I have tried to cover myself by saying [that] they're supposed to be or it's supposed that they can, it's assumed to be, and all that, and whenever a professor says that, that means it's false. Okay, so I used to think it was just the first one, that we were hung on dualism. And you have to remember it's not inevitable. We think dualism is somehow built into the structure of our language. I have a friend from Kenya who says that the mind-body problem can't even be stated in his native language, they don't have words . . . oh gosh, if only we had that language. But in any case, on the other hand, ours is pretty pervasive. I once gave a lecture on this subject in Bombay, and I was on the same platform as the Dalai Lama, a great honor. And I was, well I won't say appalled, but I was taken aback when he got up and said, "We are both a mind and a body." And I thought, "Well Descartes' influence has either spread or maybe he's tapping into some more universal mistake that we're tempted to make." Anyway, I think this is the biggest mistake and I'm going to say more about it in a few minutes. But I want

to say just get out of that hang-up of supposing that there are these mutually exclusive realms; that because consciousness is subjective in the sense that it has this first-person mode of existence, it only exists insofar as it's actually experienced by human or animal subjects; [and] that therefore it can't be part of the ordinary physical world we all live in. That's the main message I want to get across, is that the ordinary physical world we all live in contains consciousness as one of its features, and if we're embarrassed to say it's a physical feature, well then let's just get rid of that terminology and just say it's a biological feature.

The two names that we're stuck with, materialism and dualism, seem to me inadequate, and so I invented another name under pressure. I was giving a lecture in Reno, Nevada, and some guy raised his hand and said, "What's the name of your theory?" I didn't have a name. But I thought, "Well you can't have a theory if you don't have a name," so I said, "Mine is called biological naturalism," and I invented in on the spot and now I'm stuck with it. That was on the earlier slide. But the idea is that the right level for explaining consciousness is biological. But it's just part of nature, we're not postulating anything supernatural or anything that stands outside of nature.

Okay, so that I think is our first mistake that we need to overcome, and that's the biggest one. If we could get out of our tradition that says that these are mutually exclusive, then a lot of the intellectual obstacles to getting a naturalistic account of consciousness would be removed. Notice that I'm not now defending materialism, which says really consciousness as such doesn't exist, you've got to reduce it to something else or show that it was just an illusion. And I'm certainly not defending dualism that says no, no, they're really two different realms that we live in, the mental and the physical.

Now a second mistake we've got connects with a notion of causation. And we're inclined to think that whenever A causes B, they have to be completely different phenomena, different events, that causation is always a relation between discrete events ordered in time. I want to say there are other kinds of causation than that. If you ask yourself what's the causal explanation of the fact that this rostrum exerts pressure on the floor the answer is given by gravity, there's a constant gravitational attraction. But of course gravity is not the name of an event, it's the name of a permanent fixture of the universe. And if you ask what's the causal explanation of the liquid behavior of this water I have here, I could tell you a story about how the molecules are rolling around on each other in a more or less random fashion, and that the behavior at the molecular level causally explains the behavior at the system level. Similarly the behavior at the molecular level of the molecules moving in vibratory movements in lattice structures explains the solidity of the rostrum or the solidity of the floor. I want to call your attention to that, because that is bottom-up explanation, where you explain a surface feature not by citing a preexisting event, but rather by citing the behavior of the microelements of which the system that has the surface feature is composed. And I think that is the right model for seeing the relation of neuronal processes to consciousness. Consciousness is

caused by the neuronal processes, but the form of causation is bottom-up causation, so you can have the higher-level system feature causally explained by the lower-level behavior of the elements, even though the system is made up entirely of those elements.

I noticed some of the speakers were reluctant to say *cause*. One of the ways we have a fudgingness is we say, "Well the brain gives rise to consciousness." That's a hedge. Let's come right out and say it: the brain does it, the brain causes it. It isn't something that sort of just squirts out in a sort of vague way. No, it's really going on as a real feature causally explained by the brain.

Okay, reduction. Well, how much time have I got? I mean I'm just . . . have I got another ten minutes? Who's the boss? Okay, all right, all right, here we go. Reduction. The problem with reduction is that it's almost a religious term. A lot of neurobiologists tell me science is reductionist. They say that with tremendous sense of achievement. *I learned it in graduate school* is the idea. And a lot of people object to me by saying, "Too reductionist, too reductionist, this whole account."

And so I went and looked at the literature, and the truth is nobody knows what they mean by *reduction*, or rather there are half a dozen different things that they mean. There's logical reductions and causal reductions and ontological reductions and eliminative reductions, and I'm just going to tell you a couple of distinctions you need to keep in mind. First of all, what we mostly want in the sciences is a causal explanation. Now typically when we get a causal explanation, we will make an ontological reduction on the basis of the causal reduction. Now what does that mean? Well if you know that the behavior, the liquid behavior, is causally explained by the behavior of the molecules, then you have causally reduced liquidity to molecular behavior. Analogously, if you know that the solidity is causally explained by the behavior of the molecules, then you have a causal reduction. A causal reduction tells you that the reduced entity can be causally explained by the behavior of some other entity, some lower-level entity. Now typically when we make a causal reduction like that, we make an ontological reduction. We then say, "Well, liquidity just is the behavior of the molecules, or solidity just is the vibratory movement of the molecules in lattice structures, or the color red just is a certain type of wavelength." The causal reduction typically leads to an ontological reduction. But we're reluctant to do that with consciousness. Why? I mean suppose we had a complete causal account of consciousness, we could say exactly what was causing consciousness in the brain. We would still be reluctant to make an ontological reduction, to say consciousness in nothing but . . . and then follows your favorite theory, it's neuron firings at the rate of 40 Hertz synchronized between layers 4 and 6 of the cortex and the thalamus, let's say. I mean that was one theory. It didn't work, but anyway that's the kind of theory, something like that has to work. Now we would be reluctant to make an ontological reduction, to say, "Well that's all there is to consciousness," in a way that we're not reluctant to make an ontological reduction with solidity and liquidity. What's the difference? Well I

think it's not that there's some deep metaphysical difference. I mean solidity continues to feel a certain way and liquidity continues to behave in a certain way. The difference is that in the case of consciousness, the whole point of having the concept of consciousness is to name this qualitative, subjective sequence of experiences. So even if we got a complete causal account, we would still be reluctant to go the next step and say, "Well that's all it is, it's really nothing but that," because of course it's our life, we actually live this sequence of qualitative experiences. So I think we will get a causal reduction of consciousness if we have an ideal neuroscience. That's the aim is to get a complete causal account. But the causal reduction of consciousness will not lead to an ontological reduction in the way that causal reductions typically do lead to ontological reductions because we'd lose the point of having the concept if we made the ontological reduction. We still need a vocabulary to describe these qualitative subjective features of our sentient and aware life.

Now there's another distinction you need to make in reduction, and that is between those reductions that get rid of the reduced phenomenon where you show it was an illusion, there wasn't anything there, and those that don't get rid of it but just explain what it's made of, what its basis is. So I guess we can do an eliminative reduction on sunsets. I mean sunsets are just an illusion. The Sun does not really set over Mount Tamalpais, it appears to from my house, but it doesn't really, that's just an illusion. But we don't get rid of objects by saying, "Well really they can be reduced to molecules." That's a non-eliminative reduction. Now the question that a lot of people ask is, "Well if we can do an eliminative reduction of sunsets, or rainbows, let's say, why can't we do an eliminative reduction of consciousness? Why can't we show the whole thing was just an illusion?" And the answer is this: the elimination . . . eliminative reductions rest on the distinction between reality and illusion. So you can do an eliminative reduction of the sunset because it's an illusion, an eliminative reduction of the rainbow because it's an illusion. But the interesting thing about consciousness is [this]: where consciousness is concerned, the illusion is the reality. That is, if I now have the conscious illusion that I'm conscious, then I am conscious. The way that our traditional eliminative reductions work, by showing a distinction between appearance and reality, won't work for consciousness because the appearance is the reality. That is, if it consciously seems to you that you're conscious, you don't have to worry, you are conscious. If some guy comes to you and said, "Look, we've done a study and we've shown that people who meet your profile are in fact zombies, you have no consciousness whatever," you don't have to worry, you don't feel, "Oh, you know, maybe those guys are right." Don't be intimidated.

All right, so I resist the temptation to talk some more about reduction. I've talked longer than I meant to, so I'm going to skip identity. Okay, identity is another philosopher's favorite, and I'm going to go [to] my last slide, as lecturers always announce and everybody feels relieved. Unfortunately it's got a helluva lot of stuff on it, and it was really because of that that I asked for this paper to be handed out. And I didn't even ask for the slide but now we got it. Okay.

Traditional Distinctions Between Mental and Physical

Now I just decided for fun, let's make a list of our tradition. What is it that makes the mental so different from the physical? Why can't we just face the obvious fact that the mind is an ordinary part of the biological world like digestion or photosynthesis or the secretion of bile, it's just a normal part of our life? And the answer is we've got this tradition that says they're radically different. Mental is subjective, physical is objective. The mental is qualitative, the physical is quantitative. The mental has a first-person mode of existence. That means it only exists insofar as it's experienced by some eye, some self, that has it; whereas the physical has this third person or objective mode of existence. It exists apart from anybody's thoughts and feelings. Furthermore, the mental has intentionality. That's a fancy word that means aboutness or the directedness. My mental states can be about other things, so I can now be thinking about George W. Bush even though he's in Washington and I'm in New York, and physical things don't have that feature. I mean words can do that, but that's only because we use them, we impose intentionality on them. So the physical is non-intentional, and then we have this tradition that says that mental states—we get this from Descartes—are not spatially located, and they're not extended in space, whereas everything physical is spatially located and in general spatially extended. We think that the mental is not explainable by physical processes, where we think the physical world has to be causally explainable by microphysics. And then of course, we think of the mental world, at least some of us do, as somehow incapable of acting causally on the physical because they're in these two different realms, whereas the physical world we think is a causally-closed system. If the mental really existed, it couldn't affect the physical world because the physical universe is causally closed and nothing nonphysical can ever come into it.

Now I want to say that chart embodies some of the most absolutely fundamental mistakes of our civilization, and I want to chop it right in half right now. And it's this: I want to say start with the mental and grant all those stuff, that it really is subjective, I really do have these qualitative, subjective feelings. They have a first-person mode of existence and they have intentionality. But now scrap the bottom, and move the mental . . . I've even got one of these weapons that they all use, here we go, watch this. Well, I want you to notice . . . there we go, all right. Scrap it right here, move all of this over to here and just say the physical world happens to have features that are both ontologically objective and subjective. Some of them have a qualitative feel to it and a first-person mode of existence and intentionality. They're products of certain neurobiological processes. But notice, as we've been seeing for the past two days, they're spatially located and spatially extended. You don't believe it, we'll turn on our fMRI and I'll show you exactly where they are, and we're going to find out more, we've even got single-cell recordings. Furthermore, we're going to causally explain them, as we causally explain everything else, by showing that they're the result of bottom-up forms of causal explanation, and we're going to show how they act causally. And I want to say a little bit about that.

I said, with perhaps too much self-confidence, that my decision to raise my arm causes my arm to go up, but of course there is a story to be told about the activation of the motor cortex. We know the neurotransmitter, it's acetylcholine. We know how it goes to the axon end plates of the motor neurons, and there's a whole long story to be told about the cytoplasm of the muscle fiber and the actin filaments and the myosin filaments. Now why isn't that story enough? That is, you know a microstory, why doesn't that tell a story and the conscious decision to raise your arm just goes along for a kind of free ride, it doesn't really do any work. There's a name of that, it's called epiphenomenalism. That says the mind is there but it doesn't do anything, it's like a froth on a wave, it doesn't perform any work. I want to say the way we should think of it is like the car engine. It's true that there is a story to be told about the oxidization of the hydrocarbon molecules and the impact of them on the metal-alloy molecules of the piston, but when you go to your car mechanic, you don't talk about that, you say, "The damn thing won't start." What you don't say is "Look the passage of electrons between the electrodes is insufficient to oxidize the hydrocarbon molecules to the extent that the oxidization becomes self-sufficient." You just say, "The damn thing won't start, I think it's in the plugs."

Now the point is it's not that there are two different domains being described; it's the same domain from beginning to end, just different levels of description. In Berkeley you've got a lot of unemployed physicists working as car mechanics, and you might tell them that story about the oxidization of the hydrocarbons, but for normal human mechanics, you don't have to go to the micro level. And it isn't that there's some metaphysical problem about how can the spark plug ever work when really the only thing doing the job is the passage of the electrons between the electrodes; it's the same system being described at two different levels. And I want to say that's how it is when you raise your arm, it's the same system being described at different levels.

All right. Well let me just summarize, and I've only really begun this talk, but I think I got across the main message I want to get across, and that is if we can overcome certain traditional errors, if we can overcome the mistake of supposing that we live in two different realms, then we can accept consciousness on its own terms, and begin to investigate, and now the investigation is well underway, how exactly it works in the brain. There are two different research proposals. I sort of hope the other guys win because it's an easier research project, that is, what I call the building-block approach. I think the unified-field approach is pretty tough—you've got to figure out how massive rates of neuron firings in big chunks of the brain, presumably the thalamocortical system, how they cause the system to be consciousness, and that seems to me a much harder research project.

But I'm delighted with the way this conference has gone, and as several speakers have remarked, it would have been unthinkable 25 years ago. I remember when I first got interested in this, I thought, "Well, you know, I'll go talk to these medical

scientists in San Francisco," and there are a lot of people in medical school interested in the brain, and I went over and I talked to a famous neuroscientist, "Well why don't you guys get busy and solve the problem of consciousness?" And his answer after much discussion was this, he said, "Look, it's okay in my discipline to be interested in consciousness, but get tenure first, get tenure first." And I think if I had to describe the intellectual revolution that has gone on, and this is the kind of thing that drives intellectual revolutions, you can now do serious work on consciousness without tenure.

Thank you very much.

Question and Answer

David Cohen: Just one question, because we're running very late.

Man: Thanks. You have to get up fast if you want to ask questions here. You've given a very excellent but also very loose definition of consciousness, and throughout it I was struck by some of the apparent similarities between the way you were describing consciousness and things you might describe in machines and computers. For instance, you referred to the guy who wakes up in the dark room doesn't have any awareness, but he's nonetheless active. I turn my laptop on, and although there are no programs running, the sucker's on. Given that and certain other comments, I was wondering if you could comment first on the possibilities of true consciousness arising in machines. And then second—and this is perhaps more in relation to some of the researchers who've been up here—what might we eventually be able to learn about our own consciousness by sort of watching this accelerated evolution of intelligence in machines?

John Searle: Okay, could everybody hear the question? The question was, What about the prospect of machine consciousness, and how much can we learn from the progress of machine intelligence?

Now along with the dualism of the mind and the body, it seems to me another colossal mistake we make, and again it goes back to the seventeenth century, is this opposition between humans and nature and between humans and machines. If by *machine* you mean a physical system capable of performing certain functions, then we are machines. It seems to me there's no question we are machines, and that we've got these submachines, like our heart and our liver and so on. So there isn't a question about machine consciousness, I'm it and so are you.

And now the question is, Well how about an artificial machine, couldn't we build a conscious machine? Now I think we ought to hear that question like the question can you build an artificial heart that does what the heart does? And we know the answer to that. The way we got the answer to it was by figuring out how the heart works, and then building an artifact that would do the same thing. We don't know how the brain works, we don't know how the brain produces consciousness, so

until that, we don't know how to build an artifact, an artificial conscious machine. Now the problem with the notion of machine intelligence is that the machines we're talking about aren't actually intelligent—that's just a metaphor, that's an observer-relative ascription that we make to them. You see, when I do addition, I'm not very good at it but I can do three plus five equals eight. Now when my pocket calculator, I punch in three plus five and it doesn't think, "Ah, that's a three and that's a five, I've got to print out an eight," it doesn't think that at all because it doesn't think anything. It's just a hunk of junk. It's just an electronic circuit that we've programmed. Any intelligence in the computer or in the pocket calculator is entirely in the eye of the beholder, it is entirely observer-relative.

When Deep Blue beat Gary Kasparov, I was besieged by reporters. Fortunately I was in Europe and they weren't willing to spend much money on long distance, but they wanted to know isn't this a blow to human dignity, doesn't this show that the machines are really taking over? You know, you can imagine the questions that one would be asked. But the answer is, of course, it's no more a blow to human dignity than the fact that any pocket calculator can outperform any mathematician in the world. We've designed this hunk of junk to do this kind of stuff and it's terrific, but nobody should think it's of any psychological relevance. Deep Blue didn't beat Gary Kasparov in any ordinary sense because it didn't play chess, it didn't know that it was winning or losing, it didn't know that this was a pawn or this was a knight. I actually did some research on this and found out how it worked, and it is a terrific technological achievement. And as you all know, the problem in chess is the exponential problem, where you just get too many things. Well, Deep Blue could calculate 200 million moves in a fraction of a second. And that's of no psychological relevance. But Deep Blue doesn't know that it's playing chess, it doesn't know that this is a chess game, it doesn't even know that these are numbers. It's just a fancy electronic circuit without any consciousness or mental life at all. And that's a general model for so-called machine intelligence.

Now what's wrong with computers? Well the answer is—and this is what a lot of people don't get—it isn't that the computer is too much of a machine to be conscious, it's not enough of a machine. Because you see, our brain really is a machine, its operations are defined in terms of energy transfer. Computation doesn't name a machine process, it names an abstract, formal, model-theoretic, algorithmic process that we've found ways to implement in machines. But computation is purely abstract, and we put this by saying it's all syntactical, and the syntax by itself, the zeroes and ones by themselves, don't carry any causal powers. We implement that in machines, so the thing you buy in the store is a machine, but the actual computational processes are not machine processes. What's going on in here, however, are actual machine processes.