Available online at www.sciencedirect.com
ScienceDirect
Journal of Physiology - Paris 101 (2007) 169–178
Journal of Physiology Paris
www.elsevier.com/locate/jphysparis

Dualism revisited

John R. Searle *

Department of Philosophy, University of California, Berkeley, Berkeley, 94720 CA, USA

Abstract

The problem of consciousness should eventually receive a scientific solution, but there are a number of scientific and philosophical obstacles along the way. I offer solutions to the philosophical problems and proposals for approaching the scientific problems.

Keywords: Consciousness; Dualism; Materialism; Biological naturalism

Some philosophical problems, but unfortunately not very many, can receive a scientific solution. I believe one of these is the problem of consciousness. The central part of the problem can be stated quite simply: How exactly are conscious states caused by brain processes and how exactly are they realized in the brain? The two key phrases here are "caused by" and "realized in" and I will say more about these later. In the history of philosophy, this has been the center of the traditional mind–body problem: How exactly does consciousness relate to the brain and to the rest of the physical world? It seems to me in such cases, where a scientific solution is at least possible, the philosophical task is to prepare the problem conceptually, to get it into a kind of shape where it admits of being treated as a scientific problem. Specifically, a large part of the philosophical task is to clarify the problem conceptually to the point that it admits of experimental testing. You have to know what you are testing for and what counts as a positive or a negative result of the test. I think, in short, that once the problem is cleaned up, the philosophical job is over and the factual empirical issues should be solved by lab scientists. I stuck by these principles and some years ago stopped working on the problem of consciousness. I felt I had said what I wanted to say, and to my satisfaction at least I had stated the problem so that it admitted of a scientific solution. I went to work on other problems such as the structure of society (Searle, 1995) and the nature of rationality (Searle, 2002).

The difficulty, however, is that the philosophical problems won't go away so easily, and as soon as the solution to one is widely accepted a new one crops up. When I first started working in this area, the mainstream views were versions of "materialism", usually "reductionist", but sometimes "eliminativist". The idea was that a scientific account of consciousness would have to reduce it to something else, such as computation, or would eliminate it by showing that it does not really exist but is some kind of illusion. The idea of the materialists, to put it crudely, was that if consciousness really exists it must really be something else, because if it is not something else it cannot really exist. I will come back to these two versions of materialism later. Dualism, though widely held by the general public was not taken seriously by most scientists and philosophers. So most of my argumentative efforts were directed against these two versions of materialism, reductionism and eliminativism. According to reductionism, consciousness can be reduced to something else such as neuron firings or computer programs. The model here is the standard model of reductionism in science such as for example the reduction of physical objects to collections of molecules. According to eliminativism, consciousness can be eliminated as an illusion. It does not really exist, but is a common illusion. The models here are such phenomena as rainbows and sunsets. The rainbow, for example, is not really an arc in the sky, but is rather an illusion produced by the refraction

* Tel.: +1 510 642 3173; fax: +1 510 642 5160.
E-mail address: jsearle@berkeley.edu

0928-4257/$ - see front matter © 2008 Published by Elsevier Ltd.
doi:10.1016/j.jphysparis.2007.11.003

of light waves through water vapor. I think that in the past couple of decades the weaknesses of reductionism and eliminativism have become apparent to most people who work in the field. However, an odd thing has happened: dualism has gradually come to seem intellectually respectable again. One of the main aims of this article is to show the incoherence of dualism. Both dualism and materialism are false, but both are trying to say something true and we need to rescue the true part from the false part.

1. The nature of consciousness

The word "consciousness" is often said to be hard to define, but if we are just talking about a common sense definition that will identify the target of the investigation, rather than a scientific definition of the sort that can only come at the end of the investigation, then it seems to me "consciousness" is not at all hard to define, and I will repeat a definition that I have given elsewhere (Searle, 2004). Consciousness consists of those subjective states of sentience or feeling or awareness that begin when we wake from a dreamless sleep and continue on throughout the day until we become unconscious again. Dreams on this definition are a form of consciousness, though of course quite different from normal waking consciousness. Consciousness, so defined, does not imply self-consciousness. It is possible to be conscious without having a higher order consciousness of one's lower order consciousness.

When working on a problem like this, it is a good idea to remind ourselves of what we know for a fact, what features of consciousness our scientific theory should be able to account for, the data we seek to explain. Of course, we might have to change our conceptions as the research progresses, but it is a good idea to be aware of what data we start with. With this objective in mind, we can say that there are four features of consciousness that any scientific theory should attempt to explain. Of course, there are lots of other features, but these four are essential. First, every conscious state is *subjective* in the sense that it only exists as experienced by a human or animal subject. For this reason consciousness has what I call a "first-person ontology". It only exists as experienced by some "I", some human or animal subject. In this respect consciousness is unlike most of the phenomena in the world, such as mountains, molecules, and tectonic plates, that have a third-person ontology. They exist regardless of whether or not any one is experiencing them.

The second feature of consciousness, so defined, is that it is *qualitative*. There is always some qualitative feel to any conscious state. Some philosophers use the notion of "qualia" to describe this qualitative feel, but I think that this word is a source of confusion because it gives us the impression that there are two kinds of conscious states, the qualitative and the non-qualitative, and that is wrong. All conscious states by definition are qualitative.

A third feature of consciousness is that our conscious states typically come to us as part of a *unified conscious field*. So for example I don't just experience the feeling of the shirt around my neck, the sound of my voice, the sight of people around me, but I experience all of those conscious states as part of one large conscious state, my whole conscious field at the moment.

I used to think that these three features, qualitativeness, subjectivity, and unity, were independent of each other, but in fact it now seems to me that they are different aspects of one composite feature, which is the essence of consciousness. There is no way that a system could have qualitative experiences, in the sense that I have tried to explain, without subjectivity and a first-person ontology in the way I described. And there is no way that a state could be qualitative and subjective without being part of a unified conscious field. You can see this if you try to imagine your present consciousness broken into fifteen pieces. You would not have one consciousness with 15 parts, you would have 15 different conscious fields. And this, incidentally, is why the Sperry and Gazzaniga (1985) studies on the split brain patients are philosophically and scientifically so interesting. It seems that, in the pathological cases where the corpus callosum has been cut, there are two conscious fields that communicate with each other only imperfectly; and in normal, non pathological cases, the two conscious fields coalesce, in the same way that in binocular vision our two visual images coalesce.

A fourth feature of consciousness, that I won't have time to explore in much detail, is probably its most important feature. Consciousness is *intentional* in the sense that typical conscious states are *about* something or *refer to* something. If I see or think about an object, then my conscious experience is directed at or about the object; it has that object as its intentional object. Not all conscious states are intentional. For example, states of undirected anxiety are not intentional. But most conscious states are intentional, and these forms of consciousness are most interesting to us, because they mediate our relations with the rest of the world through perception, action, memory, intention, belief, desire, emotions and all the rest of our mental lives.

2. How consciousness fits into the world

With consciousness so defined, here are the facts, or at least the putative facts, that we need to account for.

1. Consciousness so defined is a real phenomenon in the real world. It really exists and we cannot get rid of it by defining it away or reducing it to something else, or by pretending that it does not really exist. Now why not? Why should consciousness be irreducible in a way that, for example, colors are not irreducible? We think (or at least many people think) that colors can be reduced to light reflectances. Why could not we do a similar reduction with consciousness? In the philosophical literature, as I mentioned earlier, typically a distinction is made between eliminative and non-eliminative reductions. Eliminative reductions show that the phenomena never existed at all, but were just an illusion. Rainbows

and sunsets, so the story goes, can be reduced in an eliminative fashion. Another kind of reduction does not show that the phenomenon does not exist but that it is really something else. Thus apparently solidity of objects can be reduced to the behavior of molecules. Material objects can be reduced to their constituent molecules. But we cannot reduce consciousness in either of these ways. We cannot show that consciousness is an illusion like sunsets or rainbows because, where the very existence of consciousness is concerned, we cannot make the distinction between reality and illusion. If I consciously have the illusion that I am conscious, then I already am conscious. Traditional eliminative reductions rest on a distinction between reality and illusion, but where the existence of consciousness is concerned, the conscious illusion is itself the reality of consciousness.

We cannot do a non-eliminative reduction of consciousness to some third-person phenomena, because if we did, we would leave out the essential first-person qualitative character of conscious experiences. Because consciousness has a first-person ontology, it cannot be reduced to anything, such as neuron firings, that has a third-person ontology. Indeed, for consciousness, we cannot even make the distinction between eliminative and non-eliminative reductions. The non-eliminative reductions end up being eliminative, because they deny the essential defining characteristic of consciousness. They deny the existence of any qualitative, subjective, unified phenomena; but unified, qualitative subjectivity is the defining feature of consciousness. All reductions of consciousness to third-person phenomena end up being eliminative, because they eliminate the essential first-person ontology of consciousness.

2. All conscious states, without exception, are caused by neurobiological processes in the brain. We now have an overwhelming amount of evidence for this, and it is no use looking for some spiritual or other sort of origin of consciousness. If the brain processes are functioning right, the subject will be conscious; if not he will not be conscious. This means that everything in consciousness, from the taste of the beer, to the sound of the music, to feeling the angst of post-industrial man under late capitalism, is caused by lower level neurobiological processes in the brain. This is an amazing fact and we ought to try to appreciate its importance. The key notion is the notion of causation. All of our conscious states are caused by neuronal processes. That means that we can do a causal reduction, but not an ontological reduction of consciousness. There is nothing to the causal power of consciousness which cannot be explained by the causal power of the neuronal base. That is why consciousness does not stand apart from the rest of the world but is an ordinary part of our human and animal biology. The causal reduction, however, does not lead to an ontological reduction, because consciousness, as a first-person phenomenon, cannot be ontologically reduced to third-person phenomena.
3. Consciousness, so defined and as caused by brain processes, is entirely realized in the brain. Now what does that mean? It means that all conscious states exist in the brain as higher level features of the neuronal system. By "higher" I mean, for example, that my conscious thoughts about my grandmother cannot be a feature of a single neuron but only of a *system* of neurons. We don't yet know how big the system has to be, but we think the thalamocortical system has a special role to play in causing and realizing consciousness. The point for the present discussion is that consciousness exists as a feature of the brain without being a feature of any individual neuron or synapse by itself. Propositions two and three amount to saying that consciousness is both *caused by* and *realized in* the brain.
4. Consciousness functions causally in producing bodily movements and other physical effects in the world. Again, this is often denied. There is always some philosopher who will tell you that conscious states cannot affect the physical world. But always remember that when I decide to raise my arm the damn thing goes up. And, notice we do not say, "Well, that is the thing about the old arm. Some days she goes up and some days she doesn't go up". My arm goes up when I want it to go up, when I decide to raise it. To put this more technically, my conscious intention-in-action causes the bodily movement of my arm raising (Searle, 1983).

Why would anyone want to deny the obvious fact that consciousness functions causally in producing our behavior? The answer, as usual, is that we are in the grip of certain traditional philosophical categories. If consciousness is not reducible to third-person phenomena, then to many people it seems that consciousness is not part of the physical world. But surely, so the story goes, the physical world is, as they say, "causally closed". What does that mean? It means anything nonphysical can never have any effect on the physical. But if consciousness really exists as a subjective phenomenon and cannot be eliminated or reduced to something objective, and if for that reason it is a nonphysical phenomenon, then it seems that it cannot affect physical reality. On this understanding we have a mysterious ghost-like phenomenon in the world that cannot have any real causal effect on the physical world. And the logical consequence is, to use the jargon of the philosophers, "epiphenomenalism". Consciousness is there alright, but it doesn't make any real difference to the real physical world. It just goes along for the ride. This argument for epiphenomenalism can either be taken to show that, if my account is right, then epiphenomenalism follows, or, if you suppose, as I do, that epiphenomenalism is not true, then my account must be wrong. I want to reject both of these conclusions. I think my account is right and I think epiphenomenalism is wrong, and I am going to answer this objection later on.

The conjunction of these four claims – consciousness is real, it is caused by brain processes, it is realized in the brain, and it functions causally – constitute an approach

to the mind that I have labeled "biological naturalism". I don't set any store by labels, but it is useful to have a name that distinguishes my view from dualism, materialism, etc. The view is biological, because it says the right level for a scientific account of consciousness is the biological level, as opposed for example to the sub-atomic or cultural levels. It is naturalistic because it says that consciousness is an ordinary part of nature along with life, digestion, photosynthesis and all the rest of it.

Biological naturalism gives us a fairly clear and well defined scientific problem: figure out exactly how the brain works to cause consciousness, how consciousness is realized in the brain and how it functions causally in our behavior. The typical pattern in science has consisted of three stages. First, we find correlations. In the case of consciousness we would like to find a correlation between the conscious state and the neurobiological processes. This has come to be called the "Neuronal Correlate of Consciousness", the NCC. The second step is to check to see whether or not the correlation is a genuine causal correlation. Do the neuronal *correlates* of consciousness actually *cause* conscious states? The usual tests for causation, as applied to this problem, would be, first, can you produce consciousness in an unconscious subject by producing the NCC, and, second, can you shut down the consciousness of a conscious subject by shutting down the NCC? All of this is familiar scientific practice. The third step, and we are a long way from reaching this step, is to get a general theoretical account. We want to know not just what the correlations are and whether they are causal, but we would like to embed all of that information in a theoretical account. Why should these causes produce these effects?

This is typical of the history of the sciences. The germ theory of disease and the DNA theory of heredity are models for the type of explanatory apparatus we are looking for.

3. The failure of materialism

As I have described it, the search for a scientific theory of consciousness looks like a typical scientific research program. What are the obstacles to pursuing it? There are a number of philosophical and empirical problems. The most obvious obstacle to a scientific account of consciousness was the "materialist" attempt to deny the irreducibility of consciousness. For a long time, literally decades, research in consciousness was damaged by certain mistakes motivated by "materialism". Many philosophers and scientists felt that science could not account for anything that was essentially subjective. The argument rested on a bad syllogism that goes as follows: Science is objective. Consciousness is subjective. Therefore, there cannot be a science of consciousness. This syllogism rests on a fallacy of ambiguity. The notions of objectivity and subjectivity in our culture are ambiguous between an epistemic sense of the distinction and an ontological sense. Epistemically, the distinction is between those propositions that can be known to be true or false independently of the feelings and attitudes of observers, and those that cannot. So if I say "Rembrandt was born in 1606" that statement is epistemically objective. Its truth or falsity can be established independently of the feelings and valuations of observers. But if I say "Rembrandt was a better painter than Vermeer", well, that is, as they say, a matter of subjective opinion. I am not even sure if it is true. This epistemic distinction should not be confused with the ontological sense of the objective–subjective distinction. This sense has to do with the modes of existence of entities. Most of the things investigated by science, such things as molecules, tectonic plates and galaxies, are ontologically objective. They exist independently of anybody's experience of them. But some other things in the world that exist are ontologically subjective, such as pains and tickles and itches. Such things exist only as experienced by a human or animal subject. Now here is the point for the present discussion. Science is indeed epistemically objective. But epistemic objectivity of a mode of inquiry does not preclude such inquiry into a domain that is ontologically subjective. You can have a perfectly objective science of an ontologically subjective domain.

Philosophy and science were seriously blocked from getting a satisfactory theory of consciousness by the confusion that led people to suppose that there could not be a science of anything ontologically subjective. In philosophy, we went through behaviorism, type–type identity theory, token–token identity theory and functionalism, among others. But the worst form of this mistake was the so-called computational theory of the mind. The idea was that all there is to consciousness is having a certain sort of computer program or programs. Maybe it is a connectionist program, or maybe it is a traditional von Neumann style program; but whatever style program, it is constitutive of consciousness. As I am about to make some criticisms of computational theory of the mind, it is important for me to emphasize at the beginning that I am certainly not opposed to computer modeling in neuroscience. It seems to me absolutely crucial. It is as crucial in the study of the brain as it is in the study of digestion, for example. Furthermore, it seems to me quite clear that there are levels of description at which we can say that the brain computes. What then are the objections to the computational theory of the mind? It seems to me there are at least the following two objections. The first point is that the computational level is not sufficient by itself for consciousness, intentionality, or the rest of the paradigmatic mental phenomena. You don't, for example, produce pain by producing a computer simulation of pain. The second objection to computation is not an objection to the activity of computer modeling but a limitation on its explanatory power. The computational description is no substitute for doing the neurobiology. If you describe the vestibular ocular reflex (VOR), for example, as solving a computational problem, and you can indeed so describe it, you still have to figure out how it works in the actual neurobiology between the semi-circular canals and the

cerebellum. In neurobiology there is no substitute for neurobiological levels of description. I will consider each of these points in order.

I demonstrated years ago with the so-called Chinese Room Argument (Searle, 1980) that the implementation of the computer program is not by itself sufficient for consciousness or cognition. Computation is defined purely formally or syntactically, whereas minds have actual mental or semantic contents, and you cannot get from the syntactical to the semantic just by having the syntactical operations and nothing else. To put this point slightly more technically, the notion "same implemented program" defines an equivalence class that is specified independently of any specific physical realization. But such a specification necessarily leaves out the biologically specific powers of the brain to cause cognitive processes. A system, me, for example, would not acquire an understanding of Chinese just by going through the steps of a computer program that simulated the behavior of a Chinese speaker.

The second objection, I think, is just as important, but it has received less attention than the first. Basic to our understanding of nature is the distinction between those features of the world that are observer-independent, that exist regardless of what anybody thinks, and those that are observer-dependent. Observer-independent phenomena would include force, mass, gravitational attraction, and tectonic plates. Observer-dependent phenomena would include money, property, and marriage. Typically, the natural sciences are about observer-independent phenomena, and the social sciences are about observer-dependent phenomena. Now what about computation? Is it observer-independent or observer-dependent? Well, if we understand computation as Turing did, in terms of the manipulation of symbols, then obviously it is observer-dependent. Except for those few cases where an agent is actually consciously going through a computational process such as adding one plus one to get two, there are no observer-independent computational processes in nature. Something is a symbol only if we treat it or regard it or use it as a symbol, and a symbolic process such as computation only exists as such relative to an interpretation. So when we are looking for computational processes in the brain, we have to make a very clear distinction between those processes where the agent is intrinsically, in an observer-independent fashion, going through some thought processes that involve symbols, and those cases where the process can be treated as computation only in an observer-relative sense. There are many processes in the brain that we can describe in computational terms, and it is quite natural to do so. But we should not think that this is some kind of a scientific discovery that discovering a computation is like discovering a new neurotransmitter. When we decide to describe the vestibular ocular reflex as a computational process, we should not think the agent is intrinsically going through any computing. When we describe the VOR as computational, this is an observer-relative description of the same sort that we make when we give observer-relative computational descriptions of digestive processes in the stomach.

The distinction between observer-independent phenomena and observer-relative phenomena is also crucial for the use of certain other concepts in cognitive science. I just argued that computation has both an observer-independent and an observer-relative sense, but the same ambiguity afflicts both the notion of "information", and the notion of "intelligence". There is a psychologically real observer-independent sense of "information", and a psychologically real observer-independent sense of "intelligence", but in addition, there are observer-dependent senses of both of these notions. I have in my head information about how to get from Berkeley to San Jose. That information is observer-independent or intrinsic to my psychological processes. I really know the way to San Jose. The map I have in my car also has the information about how to get from Berkeley to San Jose, but in the map the information is observer-dependent. It is in the eye of the beholder. Intrinsically speaking, the map is just cellulose fibers with ink stains on it. The map contains *information* only relative to our capacity to so interpret it. Similar remarks apply to the notion of "intelligence". There is a perfectly good sense in which my present computer is much smarter, that is, more "intelligent", than the computer I had ten years ago. But, I take it, there is no psychological reality to that sense, at all. On the other hand, when we say that humans are more intelligent than some other species such as dogs, we are talking about a certain psychological reality. Questions like this become important when you ask whether or not you could build an intelligent machine. In one sense, anything that follows laws of nature can be described as if it were behaving intelligently. I will now exhibit to you an intelligent machine: I here drop my pen and as it falls to the table, it computes the function $s = \frac{1}{2} gt^2$, but I take it, there is no psychological reality to the operation of the pen.

The notion of "artificial" in "artificial intelligence" suffers from an additional ambiguity. An artificial X can be either a real X produced artificially, or it can be a simulation of an X but not a real X. Thus, for example, artificial dyes are real dyes alright, but they are produced in laboratories and factories. Artificial cream, on the other hand, is not real cream, but a simulation (or imitation) of cream. These ambiguities are quite harmless as long as you do not confuse them, as long as you do not mistake one sense of "artificial intelligence" for another. The notion of "artificial intelligence" has for decades suffered from a failure on the part of its users to distinguish between creating a *simulation* of real intelligence artificially and creating *real* intelligence artificially. But of course the idea that you can create real intelligence artificially just by running a computer program runs afoul of the point I made earlier: The implementation of the program is *not by itself* sufficient for consciousness in particular or cognition in general.

In the early days artificial intelligence was defined as The effort to produce machines that behave in a way that in humans we would describe as intelligent. But the problem with that definition is that just about anything, and consequently any machine, has some level of description where we could describe it as artificial intelligence in that sense. So, for example, if I am running down a dry stream bed trying to find the line of least resistance it is going to take a lot of concentration and indeed intelligence on my part. But let a "machine", namely water, into the stream bed and it immediately finds the line of least resistance, and thus behaves "intelligently". I hope it is obvious that this observer-relative sense of intelligent behavior is of little or no psychological interest. Similarly with computation. Just about anything has some level of description where it can be described as performing computations and this is because just about anything can be described as following natural laws.

I have spent so much time on the computational theory because it was the best developed and most ambitious of the materialist theories. I now want to turn to examining some of the difficulties with current neurobiological research.

4. Difficulties in neurobiological research on consciousness

If I am right that solving the problem of consciousness ought to be a three step process – find the NCC's, find out if the correlation is a causal correlation, and get a theory in which the empirical data can be embedded – then why is progress so slow? Why don't these researchers get busy and figure it out? There are several practical reasons for the slowness of the research. One is that the brain is extremely complicated and we have very crude methods for investigating it. Each new invention, such as CAT scans and PET scans and now *f*MRI, is hailed as a great breakthrough, but the results are still rather limited. Earlier text books used to say that the problem of the research methods is that they were "invasive". What they mean by "invasive" is that you either have to kill the animal or make a mess of its brain to investigate it.

But there is I think a deeper theoretical reason why progress has been slow and that is that much of the research may be barking up the wrong tree. I distinguish for these purposes, between what I call the "unified field" conception of consciousness and the "building block" model of consciousness (Searle, 2002). The unified field conception takes seriously the view that all our conscious states come to us as part of a unified conscious field. The "building block" model of consciousness treats the conscious field as made up of a set of distinct experiential units such as, for example, seeing red, tasting wine or hearing the sound of middle C. Most of the research I know is based on the building block model. You try to follow out the path of the stimulus until it produces a conscious experience in the brain. Christof Koch's excellent book (Koch, 2004) describes this procedure very well. And there are some remarkable experiments done on the building block model.

I will discuss three of these types of experiments very briefly.[1] First, binocular rivalry. If you show the left eye a set of parallel lines and the right eye a set of horizontal lines, the subject typically does not see a grid, rather the visual system switches back and forth between producing a visual experience of parallel lines and a visual experience of horizontal lines. There is a rivalry between the two stimulus pathways, and first, one side wins and then the other side wins. Now, if we could find the exact point in the brain where the horizontal lines triumph over the vertical lines, or where the vertical lines triumph over the horizontal lines, it would seem that we have found the NCC for that particular experience.

Similar considerations apply to Gestalt switching. If you consider Wittgenstein's famous Duck–Rabbit example (Wittgenstein, 1953) (originally due to Jastrow), the stimulus is held constant, yet now I see a duck, now I see a rabbit. Once again, if we could find the point in the brain where the perception switches from that of the duck to that of the rabbit, it looks like we would have found the NCC for either the perception of the rabbit or the perception of the duck. In both of these experiments the stimulus is held constant, yet the same stimulus produces different experiences. We want to know exactly where in the brain the experience occurs, and then exactly how the brain determines that you will have one experience and not the other.

A third sort of experiment, equally famous, is the research done by Weiskrantz (1986) and others on blindsight. There are patients with damage to visual area 1 (V1) of the visual cortex which is such that they are unable to see in a certain portion of their visual field. Nonetheless, in some experiments the subject can report events occurring in the blind portion of his visual field even though he reports no visual experience of the event. So it seems there must be some pathway in the brain that is producing conscious experience, and another pathway in the brain that is giving the subject visual intentionality which is unconscious. If we could find the distinctive features of the conscious as opposed to the unconscious pathway, it looks like, once again, we might have the NCC for that sort of consciousness.

Finally, of course, the most natural way of pursuing the building block model is simply to follow out the stimulus, say, the stimulus of the color red, until it actually produces the perception, the experience of red (Koch, 2004).

I am very enthusiastic about all these lines of research but I fear they may be making a fundamental mistake: In all of these cases, the subject is already conscious. That is, it is only a conscious subject who can experience binocular rivalry, the Gestalt switch, blind-sight, or the experi-

[1] For more detailed discussion together with references see (Searle, 2002).

ence of red. And what we really need to know is: How does the brain create the conscious field in the first place? I believe we should think of perception as not so much creating consciousness, but as modifying the pre-existing conscious field.

You do not require perception to be conscious. Imagine yourself in an absolutely dark room with no sounds coming in. You wake up from a dreamless sleep and you are in a state of total alertness with only very minimal perceptual experiences. You might, for example, experience the weight of the bed clothes on you and the weight of your body against the bed. But besides these very marginal perceptual experiences, you are totally awake and alert without any perceptions. Then you get up and turn on the light and walk around and get dressed, brush your teeth, etc. Are you creating consciousness? Well in one sense you are because you now have conscious experiences that did not exist before. But I want to suggest that if we take the metaphor of the field seriously, we should think of these input stimuli as doing something like creating bumps and hillocks and valleys in the conscious field. We do not populate an empty room with our perceptions, but rather we make modifications to the conscious field that already existed. We should think of perception not as creating consciousness but as modifying a pre-existing conscious field. Now this inclines me to think that the line of attack that is most likely to succeed is to figure out how the brain creates the unified conscious field in the first place. I realize this line of research is likely to be more difficult, because you cannot rely on single cell recordings and *f*MRI in the way that you can on the "building block" model.

5. What is wrong with dualism?

It is now time to try to say what exactly is wrong with dualism. I have already said that consciousness is not ontologically reducible to brain processes. Isn't that already a kind of dualism? Isn't the irreducibility of consciousness all that dualism amounts to?

It is important in answering this question to remind ourselves that I said that consciousness was causally reducible but not ontologically reducible to neuronal processes. It is causally reducible, because there is nothing going on which cannot be causally accounted for by neuronal processes. But the causal reduction does not lead to an ontological reduction because consciousness has a first-person or subjective ontology, and for that reason cannot be reduced to something that has a third-person or objective ontology. If you try to make the reduction you leave something out, namely the subjectivity of consciousness. But once we accept the ontological irreducibility of consciousness, why can't we just say that that is a form of dualism and then live with it? What is so bad about dualism?

The real objection to dualism is that we cannot give a coherent account of reality on dualist assumptions. We cannot give an account of reality which makes a part of the real world – our conscious states – cohere with our account of the rest of the real world. Dualism postulates two distinct domains, but on this postulation it becomes impossible to explain the relationship between the domains. This incoherence has a number of consequences. Perhaps most famously, it becomes difficult, if not impossible, to explain how brain processes in one ontological domain could cause consciousness in another ontological domain. Right now I want to focus on another absurd consequence that I mentioned earlier: Epiphenomenalism. If consciousness has the features of qualitativeness, subjectivity, unity, and intentionality, but is not a part of the material or physical world, then how on earth could it possibly function causally in the physical world? On standard versions of dualism it is hard, if not impossible, to see how consciousness could have any causal impact in the world, yet we know that it does have a causal impact: I decide to raise my arm, I form a conscious intention-in-action to raise my arm, and then the arm goes up. There isn't any doubt that my conscious intention causes the arm to go up.

Suppose in answering this question further we make a chart of the world according to dualism, where we list the features of consciousness on the left side and the features of the material world, with which consciousness is supposed to be contrasted, on the right.

Consciousness	Material world
Ontological subjectivity	Ontological objectivity
Qualitativeness	Quantitative measurability
Intentionality	No intrinsic intentionality
Not spatially located/extended	Has spatial location(s)/dimensions
None of force, mass, etc.	Force, mass, gravitational attraction and electrical charge, etc.

There are lots of other features of the physical world that are supposed to be contrasted with consciousness. On the right hand side of our chart, we have spatial locations and spatial dimensions, whereas consciousness is not spatially located and not extended spatially. Furthermore, on the right hand side we have force, mass, gravitational attraction and electrical charge. Presumably none of those are supposed to exist on the left hand side. Now, the incoherence is quite obvious: Something that is not spatially located and has no physical, chemical, etc. powers could never be capable of moving my arm. But we know that my conscious intention-in-action does cause the movement of my arm. So what is the solution to this puzzle? I think the solution is obvious: Move the ontologically subjective features on the left hand side over to the right hand side, so that among the "physical", "material" features of the world we now list subjectivity, qualitativeness and intentionality along with force, mass, gravitational attraction, electrical charge, solidity, liquidity, and a very large number of others. But we will find it embarrassing to say that subjectivity etc., are "physical" or "material" because the

terminology of "physical" and "material" was designed precisely to oppose these to "mental", "spiritual", etc. So, let us get rid of the terminology, let us get rid of the apparent contrast and just say that qualitativeness, subjectivity and unity are parts of the real world along with everything else.

Now how can they function causally? To go back to our question: How can they ever move my arm? We know the answer to that question and we have known it for a long time. As usual, philosophy consists in large part of reminding ourselves of what we know already, but anyway here is what we know:

A single event in my brain has one level of description where we can describe it as a conscious intention-in-action, but it also has many other levels of description. We can describe it as involving neurons firings at a certain rate, in certain neuronal architectures. It involves certain specific neurotransmitters. We know acetylcholine is one of them, and there are no doubt others. We know that there are certain electrical phenomena that go along with all of the chemical features, and all of that causes my arm to go up. "But are you saying that the conscious state itself has electrical charges and has spatial dimensions in the brain?" The answer to that question is obviously "yes". We know that conscious events have spatial locations in the brain. Indeed, with current imaging techniques, especially with *f*MRI, we are able to identify the spatial location of many conscious events, and as we discover more about their neurobiological basis, we will be able to specify more of their specific electro-chemical features.

Indeed, we can use the alleged problem of epiphenomenalism as an argument against dualism and for biological naturalism. According to biological naturalism, conscious events have a level of description where they are neurobiological events, and we can prove this with certain very simple assumptions, as follows:

Premise (1): My intention-in-action to raise my arm caused my arm to go up.

It is typical of human intentional actions that our intentions-in-action cause bodily movements. We have a lifetime of experience to support this and in any case it is not possible to seriously doubt it.

Premise (2): Anything that caused my arm to go up in that circumstance, must have certain rather specific electro-chemical properties.

We know there must have been a sequence of neuron firings going from the motor-cortex to the muscles, we know the neurotransmitter was acetylcholine, and we know the acetylcholine at the axon-endplates of the motor neurons activated the ion channels in such a way as to attack the cytoplasm of a muscle-fiber. So we can generalize this as: Anything that caused my arm to go up must have electro-chemical properties.

Now from 1 and 2 it follows that

Conclusion (3): my intention-in-action has electro-chemical properties.

And we can generalize this for a large number of cases as follows:

(1) Certain mental events cause physical events of bodily movements.
(2) Anything that causes such bodily movements in the internal bodily fashion must have electro-chemical properties.
Therefore
(3) Such mental events have electro-chemical properties.

None of this is at all mysterious. It is just a routine example of the fact that events have higher and lower levels of description. Just as I can describe the action of my engine as the firing of the spark plug with the consequent explosion of the air–fuel mixture in the cylinder, so I can also describe it at a lower level as the passage of electrons between copper electrodes, followed by the oxidization of hydro-carbon-molecules. The same event has both higher and lower levels of description.

The fact that one and the same conscious state has different levels of description, a level of description where we describe it in terms of its subjective properties, and another level of description where we describe it in terms of its chemical and electromagnetic properties should be no more mysterious to us than the fact that we describe the behavior of a car engine at different levels. The chief difference between the two cases, of course, is that the mental event has a level of description where it is ontologically subjective and that is not the case with the explosion in the cylinder of the car engine. We have already accounted for this difference by saying that the ontological subjectivity of the mental with its consequent ontological irreducibility does not imply that it is causally irreducible. On the contrary: We get a causal reduction of conscious events to neurobiological events.

I actually think philosophically or metaphysically the situation is really that simple. However, neurobiologically the whole problem is extremely complicated and we are nowhere near to getting a neurobiological solution.

We can summarize these points by saying that the reason that consciousness can function causally in a "physical" world is that it has physical properties. Every conscious state is realized in a certain physical structure in the brain and has the conscious powers of all of these physical structures. The bottom line of this discussion is that we know that among the "physical" events that occur in the world, some have subjectivity, qualitativeness and intentionality, and these events function causally along with any other event that has different levels of description.

But doesn't that leave us open to the objection that this is just the old identity theory in disguise? Aren't we just saying that conscious states are neurobiological states of the brain? Well, in one way it seems to me that so stated the identity theory is absolutely right and could hardly be

false. However, historically the identity theorists that I know, with very few exceptions, had a reductionist motive. They wanted to get rid of subjectivity. They wanted to say that consciousness is *nothing but* neurobiological states of the brain neurobiologically described in third-person terms. I have argued in this article that we know independently that that claim is false.

But doesn't that leave us with a puzzle? Why should evolution have given us consciousness in the first place? What is the evolutionary function of consciousness if we could imagine that the laws of nature were different so that the lower level processes could have their causal effects without the higher level of consciousness? This is a weird question and it is important to see the strange presuppositions that underlie it. The crucial part of it is that we have to imagine that the laws of nature are different. That is, we have to imagine that the world is radically different in such a way that I can intentionally raise my arm without any mental life at all. But of course in the real world, consciousness performs a whole family of evolutionary functions. It enables us to coordinate and manipulate an enormous volume of information simultaneously, consecutively and coherently. So I can imagine that the world can function without consciousness, if I imagine suitable changes in the laws of nature. But that is a bit like saying that I can imagine birds flying without wings, if I imagine that evolution equipped them all with rocket engines and a continuous supply of rocket fuel. Yes, if the laws of nature were suitably different all kinds of evolutionary functions would not be performed at all or would be performed by different evolutionary mechanisms. But in the world we actually live in, the only way a beast like me is going to behave intelligently and cope with the world in the way that I do is if it is conscious at least some of the time. It is true that lots of my mental phenomena are unconscious, but we understand them as "mental" precisely because we understand them as the kind of thing that could be conscious even if in a specific case they are not conscious.

6. Refutation of the argument for dualism

There is a standard argument for dualism that, I think, has become the most popular argument against materialism. Indeed, in a weaker and more cautious version I have used this argument myself. Here is how it goes. Take any version of materialism, either the identity theory that says the mind is nothing but the brain neurobiologically described, or the computational theory, or functionalism, or pick your favorite materialist theory. Now you could imagine a circumstance in which all of the material predicates were satisfied by a system, but none of the mental predicates were. That is, you might imagine a system that looked just like me and behaved just like me, but was unconscious. This is sometimes called the "zombie argument" because it would be possible for there to be zombies that behaved just like me, but lacked any of my mental traits. Another way to put this argument is to say that the problem with all forms of materialism is they are confronted with the question: Are there two kinds of phenomena going on in there only one? And the answer has to be: Where consciousness is concerned, there are first-person phenomena and third-person phenomena. Materialism is forced to say there is only one kind of thing, the third-person phenomena, but we all know from our experiences, that in addition to the neuron firings, the computer programs, the behavior, etc., there are my subjective, qualitative conscious states. So far so good. But recently the zombie argument has been extended in a way that I think is mistaken. Here is how it goes. You can imagine the whole history of the universe, exactly as it is and exactly as it has been, particle for particle, molecule for molecule, but minus consciousness. Of course you would have to imagine that the laws of nature are different in this state of affairs from what they are in the real world, but all the same, it looks like consciousness is something added on. It looks like it is some separate phenomenon.

This argument is sometimes put in the form as an imagined parable about the creation of the world. Imagine God creating the world. First he has to create all the physical particles. Then he has to add the laws that determine the behavior of the physical particles. And finally, after He has done all that, He still has to add consciousness. He might have to add some more laws to get consciousness, but consciousness is something in addition to the physical particles and physical behavior.

At the most fundamental level I think this argument is mistaken. And the reason is, at the most fundamental level, the laws don't just describe the behavior of independently existing physical particles, rather satisfying the laws is partly constitutive of the particles. So, for example, God does not first create electrons and then decide to give them a negative charge, rather the negative charge is part of the essence of being an electron. The picture, in short, that underlies the extended zombie argument is this: a lot of philosophers like to think of the world as made up of physical particles where the physical particles are like tiny grains of sand and then the laws of nature determine which ways the grains of sand are blowing. On this conception, once you have the grains of sand and you have the laws of nature that determine which way they are blowing, you still have to add consciousness to the grains of sand. I think there is something fishy about this picture. I know that a lot of philosophers have this picture and even some physicists talk this way, but I think it is incorrect to think that the particles exist independently of the laws that determine their behavior. Rather, the laws partly define the particles: The laws are partly constitutive of the particles. On my view, given the constitution of reality, consciousness has to follow in the same way that any other biological property, such as mitosis, meiosis, photosynthesis, digestion, lactation, or the secretion of bile, follows.

7. Conclusion

I said earlier that both materialism and dualism are trying to say something true and we need to rescue the truth from the urge to falsehood. Materialism tries to say truly that the universe is entirely material, that it is made up entirely of physical particles in fields of force. And dualism tries to say truly that subjective, qualitative states of consciousness cannot be reduced or eliminated in terms of any third-person phenomena. Materialism says falsely that consciousness, as irreducible, qualitative subjectivity, does not exist. Dualism says falsely that irreducible, subjective, qualitative phenomena are something in addition to, something over and above, the real, physical world. Now I think we can see how to say what is true in both without saying the false part. The world does indeed consist of physical particles in fields of force (or whatever the ultimate entities of the true physics turn out to be). But among the higher level features of these, entirely caused by the behavior of the lower level elements, are consciousness with its qualitative subjectivity. These are causally, but not ontologically, reducible to the behavior of the neuronal substrates. But they are part of the ordinary physical world like any other biological phenomenon.

References

Gazzaniga, M., 1985. The Social Brain. Basic Books, New York.

Koch, Christof, 2004. The Quest for Consciousness. Roberts & Company Publishers, Englewood/Colorado.

Searle, J.R., 1980. Minds, Brains and Programs. In: Behavioral and Brain Sciences, vol. 3. Cambridge University Press, Cambridge, pp. 417–457.

For an explanation of intentions-in-action and how they relate to intentionality and actions in general, see Searle, J.R., 1983. Intentionality. An Essay in the Philosophy of Mind. Cambridge University Press, Cambridge (especially chapter 3).

Searle, J.R., 1995. The Construction of Social Reality. The Free Press, New York.

Searle, J.R., 2002. Rationality in Action. MIT Press, Cambridge, MA.

Searle, J.R., 2002. Consciousness, Annual Review of Neuroscience, 23/2000. Reprinted in: Consciousness and Language. Cambridge University Press, Cambridge.

Searle, J.R., 2004. Mind: A Brief Introduction. Oxford University Press, Oxford.

Weiskrantz, L., 1986. Blindsight: A Case Study. Clarendon Press, Oxford.

Wittgenstein, L., 1953. Philosophical Investigations. Blackwell, Oxford.