ELSEVIER

Brain Research Reviews 26 (1998) 379–387

BRAIN RESEARCH REVIEWS

How to study consciousness scientifically [1]

John R. Searle

University of California at Berkeley, Department of Philosophy, 148 Moses Hall, Berkeley, CA 94720-2390, USA

Keywords: Consciousness; Mind; Brain; Neurobiology

Contents

Thesis 1 380
Thesis 2 381
Thesis 3 382
Thesis 4 383
Thesis 5 383
Thesis 6 384
Thesis 7 385
Thesis 8 385
Thesis 9 386
Conclusion 387
References 387

The neurosciences have now advanced to the point that we can address — and perhaps, in the long run, even solve — the problem of consciousness as a scientific problem like any other. However there are a number of philosophical obstacles to this project. The aim of this article is to address and try to overcome some of those obstacles. Because the problem of giving an adequate account of consciousness is a modern descendant of the traditional 'Mind–Body Problem', I will begin with a brief discussion of the traditional problem.

The mind–body problem can be divided into two problems, the first is easy to solve, the second is much more difficult. The first is this: What is the *general* character of the relations between consciousness and other mental phenomena on the one hand and the brain on the other. The solution to the easy problem can be given with two principles: First, consciousness and indeed all mental phenomena are *caused by lower level neurobiological processes in the brain*; and, second, consciousness and other mental phenomena are *higher level features of the brain*. I have expounded this solution to the mind–body problem in a number of writings, so I won't say more about it here (See for example [1,2]).

The second and more difficult problem is to explain in detail how it actually works in the brain. Indeed I believe that a solution to the second problem would be the most important scientific discovery of the present era. When — and if — it is made it will be an answer to this question: ''How exactly do neurobiological processes in the brain

[1] Published on the World Wide Web on 24 November 1998.

cause consciousness?'' Given our present models of brain functioning it would be an answer to the question, ''How exactly do the lower-level neuronal firings at synapses cause all of the enormous variety of our conscious (subjective, sentient, aware) experiences?'' Perhaps we are wrong to think that neurons and synapses are the right anatomical units to account for consciousness, but we do know that some elements of brain anatomy must be the right level of description for answering our question. We know that because we know that brains do cause consciousness, in a way that elbows, livers, television sets, cars and commercial computers do not do it, and therefore the special features of brains, features that they do not have in common with elbows, livers, etc., must be essential to the causal explanation of consciousness.

The explanation of consciousness is essential for explaining most of the features of our mental life because in one way or another they involve consciousness. How exactly do we have visual and other sorts of perceptions? What exactly is the neurobiological basis of memory, and of learning? What are the mechanisms by which nervous systems produce sensations of pain? What, neurobiologically speaking, are dreams and why do we have them? Even: why does alcohol make us drunk and why does bad news make us feel depressed? In fact I do not believe we can have an adequate understanding of *unconscious* mental states until we know more about the neurobiology of consciousness.

As I said at the beginning, our ability to get an explanation of consciousness — a precise neurobiology of consciousness — is in part impeded by a series of philosophical confusions. This is one of those areas of science, (and they are actually more common than you might suppose) where scientific progress is blocked by philosophical error. And since many scientists and philosophers make these errors, I am going to devote this article to trying to remove what I believe are some of the most serious philosophical obstacles to understanding the relation of consciousness to the brain.

Since it will seem presumptuous for a philosopher to try to advise scientists in an area outside his special competence, I want to begin by making a few remarks about the relation of philosophy to science and about the nature of the problem we are discussing. 'Philosophy' and 'science' do not name distinct subject matters in the way that 'molecular biology', 'geology', and 'the history of Renaissance painting' name distinct subject areas; rather at the abstract level at which I am now considering these issues, there is no distinction of subject matter because, in principle at least, both are universal in subject matter. And of the various parts of this universal subject matter, each aims for knowledge. When knowledge becomes systematic we are more inclined to call it scientific knowledge, but knowledge as such contains no restriction on subject matter. 'Philosophy' is in large part the name for all those questions which we do not know how to answer in the systematic way that is characteristic of science. These questions include, but are not confined to, the large family of conceptual questions that have traditionally occupied philosophers: What is truth, justice, knowledge, meaning, etc. For the purposes of this discussion the only important distinction between philosophy and science is this: Science is systematic knowledge; philosophy is in part an attempt to get us to the point where we can have systematic knowledge. This is why science is always right and philosophy is always wrong: as soon as we think we really know something we stop calling it philosophy and start calling it science. Beginning in the seventeenth century the area of systematic knowledge, i.e. scientific knowledge, increased with the growth of systematic methods for acquiring knowledge. Unfortunately most of the questions that most bother us have not yet been amenable to the methods of scientific investigation. But we do not know how far we can go with those methods and we should be reluctant to say a priori that such and such questions are beyond the reach of science. I will have more to say about this issue later, because many scientists and philosophers think that the whole subject of consciousness is somehow beyond the reach of science.

A consequence of these points is that there are no 'experts' in philosophy in the way that there are in the sciences. There are experts on the history of philosophy and experts in certain specialized corners of philosophy such as mathematical logic, but on most of the central philosophical questions there is no such thing as an established core of expert opinion. I remark on this because I frequently encounter scientists who want to know what philosophers think about a particular issue. They ask these questions in a way that suggests that they think there is a body of expert opinion that they hope to consult. But in the way that there is an answer to the question, ''What do neurobiologists currently think about LTP (long term potentiation)?''; there is no comparable answer to the question, ''What do philosophers currently think about consciousness?'' Another consequence of these points is that you have to judge for yourself whether what I have to say in this article is true. I cannot appeal to a body of expert opinion to back me up. If I am right, what I say should seem obviously true, once I have said it and once you have thought about it.

The method I will use in my attempt to clear the ground of various philosophical obstacles to the examination of the question, ''How exactly do brain processes cause consciousness?'' is to present a series of views that I think are false or confused and then, one by one, try to correct them by explaining why I think they are false or confused. In each case I will discuss views I have found to be widespread among practicing scientists and philosophers.

Thesis 1

Consciousness is not a suitable subject for scientific investigation because the very notion is ill defined. We do

not have anything like a scientifically acceptable definition of consciousness and it is not easy to see how we could get one, since consciousness is unobservable. The whole notion of consciousness is at best confused and at worst it is mystical.

Answer to Thesis 1.

We need to distinguish analytic definitions, which attempt to tell us the essence of a concept, from common sense definitions, which just make clear what we are talking about. An example of an analytic definition is

Water = df.H_2O

A common sense definition of the same word is, for example,

Water is a clear, colorless, tasteless liquid. It falls from the sky in the form of rain, and it is the liquid which is found in lakes, rivers and seas.

Notice that analytic definitions typically come at the *end*, not at the beginning of a scientific investigation. What we need at this point in our work is a common sense definition of consciousness and such a definition is not hard to give: 'Consciousness' refers to those states of *sentience* or *awareness* that typically begin when we wake from a dreamless sleep and continue through the day until we fall asleep again, die, go into a coma or otherwise become 'unconscious'. Dreams are also a form of consciousness, though in many respects they are quite unlike normal waking states.

Such a definition, whose job is to identify the target of scientific investigation and not to provide an analysis, is adequate and indeed is exactly what we need to begin our study. Because it is important to be clear about the target, I want to note several consequences of the definition:

First, consciousness, so defined, is an inner qualitative, subjective state typically present in humans and the higher mammals. We do not at present know how far down the phylogenetic scale it goes, and until we get an adequate scientific account of consciousness it is not useful to worry about whether, e.g. snails are conscious.

Second, consciousness so defined should not be confused with *attention* because in this sense of consciousness there are many things I am conscious of that I am not paying attention to, such as the feeling of the shirt on my back for example.

Third, consciousness so-defined should not be confused with self-consciousness. Consciousness, as I am using the word, refers to any state of sentience or awareness, but self-consciousness, in which the subject is aware of himself or herself, is a very special form of consciousness, perhaps peculiar to humans and the higher animals. Forms of consciousness such as feeling a pain do not necessarily involve a consciousness of a self as a self.

Fourth, I experience my own conscious states, but I can neither experience nor observe those of another human or animal, nor can they experience or observe mine. But the fact that the consciousness of others is 'unobservable' does not by itself prevent us from getting a scientific account of consciousness. Electrons, black holes and the Big Bang are not observable by anybody, but that does not prevent their scientific investigation.

Thesis 2

Science is, by definition, ***objective****, but on the definition of consciousness you have provided you admit it is* ***subjective****. So, it follows from your definition that there cannot be a science of consciousness.*

Answer to Thesis 2.

I believe that this statement reflects several centuries of confusion about the distinction between objectivity and subjectivity. It would be a fascinating exercise in intellectual history to trace the vicissitudes of the objective/subjective distinction. In Descartes's writings in the seventeenth century, 'objective' had something close to the opposite of its current meaning [3]. Sometime — I don't know when — between the seventeenth century and the present, the objective–subjective distinction rolled over in bed.

However, for present purposes, we need to distinguish between the epistemic sense of the objective–subjective distinction and the ontological sense. In the epistemic sense, objective claims are objectively verifiable or objectively knowable, in the sense that they can be known to be true or false in a way that does not depend on the preferences, attitudes or prejudices of particular human subjects. So, if I say, for example, ''Rembrandt was born in 1606'', the truth or falsity of that statement does not depend on the particular attitudes, feelings or preferences of human subjects. It is, as they say, a matter of objectively ascertainable fact. This statement is epistemically objective. It is an objective fact that Rembrandt was born in 1606.

This statement differs from subjective claims whose truth cannot be known in this way. So, for example, if I say ''Rembrandt was a better painter than Rubens'', that claim is epistemically subjective, because, as we would say, it's a matter of subjective opinion. There is no objective test, nothing independent of the opinions, attitudes and feelings of particular human subjects, which would be sufficient to establish that Rembrandt is a better painter than Rubens.

I hope the distinction between objectivity and subjectivity in the epistemic sense is intuitively clear. But there is another distinction which is related to the *epistemic* objective–subjective distinction, but should not be confused with it and that is, the distinction between *ontological* objectivity and subjectivity. Some entities have a subjective mode of existence. Some have an objective mode of existence. So, for example, my present feeling of pain in my lower back is ontologically subjective in the sense that

it only exists as experienced by me. In this sense, all conscious states are ontologically subjective, because they have to be experienced by a human or an animal subject in order to exist. In this respect, conscious states differ from, for example, mountains, waterfalls or hydrogen atoms. Such entities have an objective mode of existence, because they do not have to be experienced by a human or animal subject in order to exist.

Given this distinction between the *ontological* sense of the objective–subjective distinction, and the *epistemic* sense of the distinction, we can see the ambiguity of the claim made in Thesis 2. Science is indeed objective in the epistemic sense. We seek truths that are independent of the feelings and attitudes of particular investigators. It doesn't matter how you feel about hydrogen, whether you like it or don't like it, hydrogen atoms have one electron. It is not a matter of opinion. That is why the claim that Rembrandt is a better painter than Rubens is not a scientific claim. But now, the fact that science seeks objectivity in the epistemic sense should not blind us to the fact that there are ontologically subjective entities that are as much a matter of scientific investigation as any other biological phenomenon. We can have epistemically objective knowledge of domains that are ontologically subjective. So, for example, in the epistemic sense, it is an objective matter of fact —not a matter of anybody's opinion— that I have pains in my lower back. But the existence of the pains themselves is ontologically subjective.

The answer, then, to Thesis 2 is that the requirement that science be objective does not prevent us from getting an epistemically objective science of a domain that is ontologically subjective.

Thesis 3

There is no way that we could ever give an intelligible causal account of how anything subjective and qualitative could be caused by anything objective and quantitative, such as neurobiological phenomena. There is no way to make an intelligible connection between objective third person phenomena, such as neuron firings and qualitative, subjective states of sentience and awareness.

Answer to Thesis 3

Of all the theses we are considering, this seems me the most challenging. In the hands of some authors, e.g., Thomas Nagel [4], it is presented as a serious obstacle to getting a scientific account of consciousness using anything like our existing scientific apparatus. The problem, according to Nagel, is that we have no idea how objective phenomena, such as neuron firings, could necessitate, could make it unavoidable, that there be subjective states of awareness. Our standard scientific explanations have a kind of necessity, and this seems to be absent from any imaginable account of subjectivity in terms of neuron firings. What fact about neuron firings in the thalamus could make it necessary that anybody who has those firings in that area of the brain must feel a pain, for example?

However, though I think this is a serious problem for philosophical analysis, for the purpose of the present discussion, there is a rather swift answer to it: We know in fact that it happens. That is, we know as a matter of fact that brain processes cause consciousness. The fact that we don't have a theory that explains how it is possible that brain processes could cause consciousness, is a challenge for philosophers and scientists. But it is by no means a challenge to the fact that brain processes do in fact cause consciousness, because we know independently of any philosophical or scientific argument that they do. The mere fact that it happens is enough to tell us that we should be investigating the form of its happening and not challenging the possibility of its happening.

So I accept the unstated assumption behind Thesis 3: Given our present scientific paradigms it is not clear how consciousness could be caused by brain processes. But I see that as analogous to: Within the explanatory apparatus of Newtonian mechanics, it is not clear how there could exist a phenomenon such as electro-magnetism; within the explanatory apparatus of nineteenth century chemistry, it is not clear how there could be a nonvitalistic, chemical explanation of life. That is, I see the problem as analogous to earlier apparently unsolvable problems in the history of science. The challenge is to forget about how we think the world ought to work, and instead figure out how it works in fact.

My own guess — and at this stage in the history of knowledge it is only a speculation — is that when we have a general theory of how brain processes cause consciousness, our sense that it is somehow arbitrary or mysterious will disappear. In the case of the heart for example it is clear how the heart causes the pumping of blood. Our understanding of the heart is such that we see the necessity. Given these contractions blood must flow through the arteries. What we so far lack for the brain is an analogous account of how the brain causes consciousness. But if we had such an account — a general causal account — then it seems to me our sense of mystery and arbitrariness would disappear.

It is worth pointing out that our sense of mystery has already changed since the seventeenth century. To Descartes and the Cartesians, it seemed mysterious that a physical impact on our bodies should cause a sensation in our souls. But we have no trouble in sensing the necessity of pain given certain sorts of impacts on our bodies. We do not think it at all mysterious that the man whose foot is caught in the punch press is suffering terrible pain. We have moved the sense of mystery inside. It now seems mysterious to us that neuron firings in the thalamus should cause sensations of pain. And I am suggesting that a thorough-going neurobiological account of how and why exactly it happens would remove this sense of mystery.

Thesis 4

All the same, within the problem of consciousness we need to separate out the qualitative, subjective features of consciousness from the measurable objective aspect which can be properly studied scientifically. These subjective features, sometimes called 'qualia', can be safely left on one side. That is, the problem of qualia needs to be separated from the problem of consciousness. Consciousness can be defined in objective third person terms and the qualia can then be ignored. And, in fact, this is what the best neurobiologists are doing. They separate the general problem of consciousness from the special problem of qualia.

Answer to Thesis 4.

I would have not have thought that this thesis — that consciousness could be treated separately from qualia — was commonly held until I discovered it in several recent books on consciousness[5]. The basic idea is that the problem of qualia can be carved off from consciousness and treated separately or better still, simply brushed aside. This seems to me profoundly mistaken. There are not two problems, the problem of consciousness and then a subsidiary problem, the problem of qualia. *The problem of consciousness is identical with the problem of qualia, because conscious states are qualitative states right down to the ground.* Take away the qualia and there is nothing there. This is why that I seldom use the word 'qualia', except in sneer quotes, because it suggests that there is something else to consciousness besides qualia, and there isn't. Conscious states by definition are inner, qualitative, subjective states of awareness or sentience.

Of course, it is open to anybody to define these terms as he likes and use the word 'consciousness' for something else. But then we would still have the problem of what I am calling 'consciousness', which is the problem of accounting for the existence of our ontologically subjective states of awareness. The point for the present discussion is that the problem of consciousness and the problem of so called qualia is the same problem; and you cannot evade the identity by treating consciousness as some third person, ontologically objective phenomenon and setting qualia on one side, because to do so is simply to change the subject.

Thesis 5

Even if consciousness did exist, as you say it does, in the form of subjective states of awareness or sentience, all the same it couldn't make a real difference to the real physical world. It would just be some surface phenomenon that didn't matter causally to the behavior of the organism in the world. In the current philosophical jargon, consciousness would be epiphenomenal. It would be like surface reflections on the water of the lake or the froth on the wave coming to the beach. Science can offer an explanation why there are surface reflections and why the waves have a froth, but in our basic account of how the world works, these surface reflections and bit of froth are themselves caused, but are causally insignificant in producing further effects. Think of it this way: If we were doing computer models of cognition, we might have one computer that performed cognitive tasks, and another one, just like the first, except that the second computer was lit up with a purple glow. Now that is what consciousness amounts to: a scientifically irrelevant, luminous purple glow. And the proof of this point is that for any apparent explanation in terms of consciousness a more fundamental explanation can be given in terms of neurobiology. For every explanation of the form, for example, my conscious decision to raise my arm caused my arm to go up, there is a more fundamental explanation in terms of motor neurons, acetylcholine, etc.

Answer to Thesis 5.

It might turn out that in our final scientific account of the biology of conscious organisms, the consciousness of these organisms plays only a small or negligible role in their life and survival. This is logically possible in the sense, for example, that it might turn out that DNA is irrelevant to the inheritance of biological traits. It might turn out that way but it is most unlikely, given what we already know. Nothing in Thesis 5 is a valid argument in favor of the causal irrelevance of consciousness.

There are indeed different levels of causal explanation in any complex system. When I consciously raise my arm, there is a macro level of explanation in terms of conscious decisions, and a micro level of explanation in terms of synapses and neurotransmitters. But, as a perfectly general point about complex systems, the fact that the macro level features are themselves caused by the behavior of the micro elements and realized in the system composed of the micro elements does not show that the macro level features are epiphenomenal. Consider for example, the solidity of the pistons in my car engine. The solidity of a piston is entirely explainable in terms of the behavior of the molecules of the metal alloy of which the piston is composed; and for any macro level explanation of the workings of my car engine given in terms of pistons, the crank shaft, sparkplugs, etc., there will be micro levels of explanation given in terms of molecules of metal alloys, the oxidation of hydrocarbon molecules, etc. But this does not show that the solidity of the piston is epiphenomenal. On the contrary such an explanation explains why you can make effective pistons out of steel and not out of butter or papier maché. Far from showing the macro level to be epiphenomenal, the micro level of explanation explains, among other things, why the macro levels are causally efficacious. That is, in such cases the bottom up causal explanations of macro level phenomena show why the macrophenomena are not epiphenomenal. An adequate science of consciousness should analogously show how my conscious decision to raise my arm causes my arm to go up by showing how the consciousness, as a biological

feature of the brain, is grounded in the micro level neurobiological features.

The point that I am making here is quite familiar: It is basic to our world view that higher-level or macro features of the world are grounded in or implemented in micro structures. The grounding of the macro in the micro does not by itself show that the macro phenomena are epiphenomenal. Why then do we find it difficult to accept this point where consciousness and the brain are concerned? I believe the difficulty is that we are still in the grip of a residual dualism. The claim that mental states must be epiphenomenal is supported by the assumption that because consciousness is non-physical, it could not have physical effects. The whole thrust of my argument has been to reject this dualism. Consciousness is an ordinary biological, and therefore physical, feature of the organism, as much as digestion or photosynthesis. The fact that it is a physical biological feature does not prevent it from being an ontologically subjective mental feature. The fact that it is both a higher level and a mental feature is no argument at all that it is epiphenomenal, any more than any other higher level biological feature is epiphenomenal. To repeat, it might turn out to be epiphenomenal, but no valid a priori philosophical argument has been given which shows that it must turn out that way.

Thesis 6

Your last claims fail to answer the crucial question about the causal role of consciousness. That question is: What is the evolutionary function of consciousness? No satisfactory answer has ever been proposed to that question, and it is not easy to see how one will be forthcoming since it is easy to imagine beings behaving just like us who lack these 'inner, qualitative, states' you have been describing.

Answer to Thesis 6.

I find this point very commonly made, but if you think about it I hope you will agree that it is a very strange claim to make. Suppose someone asked, what is the evolutionary function of wings on birds? The obvious answer is that for most species of birds the wings enable them to fly and flying increases their genetic fitness. The matter is a little more complicated because not all winged birds are able to fly (consider penguins, for example) and more interestingly, according to some accounts, the earliest wings were really stubs sticking out of the body that functioned to help the organism keep warm. But there is no question that relative to their environments, seagulls, for example, are immensely aided by having wings with which they can fly. Now suppose somebody objected by saying that we could imagine the birds flying just as well without wings. What are we supposed to imagine? That the birds are born with rocket engines? That is, the evolutionary question only makes sense given certain background assumptions about how nature works. Given the way that nature works, the primary function of the wings of most species of birds is to enable them to fly. And the fact that we can imagine a science fiction world in which birds fly just as well without wings is really irrelevant to the evolutionary question. Now similarly with consciousness. The way that human and animal intelligence works is through consciousness. We can easily imagine a science fiction world in which unconscious zombies behave exactly as we do. Indeed, I have actually constructed such a thought experiment, to illustrate certain philosophical points about the separability of consciousness and behavior [6]. But that is irrelevant to the actual causal role of consciousness in the real world.

When we are forming a thought experiment to test the evolutionary advantage of some phenotype, what are the rules of the game? In examining the evolutionary functions of wings, no one would think it allowable to argue that wings are useless because we can imagine birds flying just as well without wings. Why is it supposed to be allowable to argue that consciousness is useless because we can imagine humans and animals behaving just as they do now but without consciousness? As a science fiction thought experiment, that is possible, but it is not an attempt to describe the actual world in which we live. In our world, the question 'What is the evolutionary function of consciousness?' is like the question, 'What is the evolutionary function of being alive?' After all, we could imagine beings who outwardly behaved much as we do but are all made of cast iron and reproduce by smelting and who are all quite dead. I believe that the standard way in which the question is asked reveals fundamental confusions. In the case of consciousness the question 'What is the evolutionary advantage of consciousness?' is asked in a tone which reveals that we are making the Cartesian mistake. We think of consciousness as not part of the ordinary physical world of wings and water, but as some mysterious nonphysical phenomenon that stands outside the world of ordinary biological reality. If we think of consciousness biologically, and if we then try to take the question seriously, the question, 'What is the evolutionary function of consciousness?' boils down to, for example: 'What is the evolutionary function of being able to walk, run, sit, eat, think, see, hear, speak a language, reproduce, raise the young, organize social groups, find food, avoid danger, raise crops, and build shelters?' *because for humans all of these activities, as well as countless others essential for our survival, are conscious activities.* That is, 'consciousness' does not name a separate phenomenon, isolable from all other aspects of life, but rather 'consciousness' names the mode in which humans and the higher animals conduct the major activities of their lives.

This is not to deny that there are interesting biological questions about the specific forms of our consciousness. For example, what evolutionary advantages, if any, do we derive from the fact that our color discriminations are conscious and our digestive discriminations in the diges-

tive tract are typically not conscious? But as a general challenge to the reality and efficacy of consciousness, the skeptical claim that consciousness serves no evolutionary function is without force.

Thesis 7

Causation is a relation between discrete events ordered in time. If it were really the case that brain processes cause conscious states, then conscious states would have to be separate events from brain processes and that result would be a form of dualism, dualism of brain and consciousness. Any attempt to postulate a causal explanation of consciousness in terms of brain processes is necessarily dualistic and therefore incoherent. The correct scientific view is to see that consciousness is ***nothing but*** *patterns of neuron firings.*

Answer to Thesis 7

This thesis expresses a common mistake about the nature of causation. Certainly there are many causal relations that fit this paradigm. So, for example, in the statement, ''the shooting caused the death of the man'', we describe a sequence of events where first the man was shot and then he died. But there are lots of causal relations that are not discrete events but are permanent causal forces operating through time. Think of gravitational attraction. It isn't the case that there is first gravitational attraction, and then, later on, the chairs and tables exert pressure against the floor. Rather, gravitational attraction is a constant operating force and, at least in these cases, the cause is cotemporal with the effect.

More importantly for the present discussion, there are many forms of causal explanation that rely on *bottom up* forms of causings. Two of my favorite examples are solidity and liquidity. This table is capable of resisting pressure and is not interpenetrated by solid objects. But of course, the table, like other solid objects, consists entirely of clouds of molecules. Now, how is it possible that these clouds of molecules exhibit the causal properties of solidity? We have a theory: Solidity is caused by the behavior of molecules. Specifically, when the molecules move in vibratory movements within lattice structures, the object is solid. Now, somebody might say ''Well, but then solidity consists in nothing but the behavior of the molecules'', and in a sense that has to be right. However, solidity and liquidity are causal properties in addition to the summation of the molecule movements. Some philosophers find it useful to use the notion of an 'emergent property.' I don't find this a very clear notion, because it is so confused in the literature. But if we are careful, we can give a clear sense to the idea that consciousness, like solidity and liquidity, is an emergent property of the behavior of the micro-elements of a system that is composed of those micro-elements. An emergent property, so defined, is a property that is explained by the behavior of the micro-elements but cannot be deduced simply from the composition and the movements of the micro-elements. In my writings, I use the notion of a 'causally emergent' property (cf. Ref. [7]) and in that sense, liquidity, solidity and consciousness are all causally emergent properties. They are emergent properties caused by the micro-elements of the system of which they are themselves features.

The point I am eager to insist on now is simply this: The fact that there is a causal relation between brain processes and conscious states does not imply a dualism of brain and consciousness any more than the fact that the causal relation between molecule movements and solidity implies a dualism of molecules and solidity. I believe the correct way to see the problem is to see that consciousness is a higher level feature of the system, the behavior of whose lower level elements cause it to have that feature.

But this claim leads to the next problem — that of reductionism.

Thesis 8

Science is by its very nature ***reductionistic****. A scientific account of consciousness must show that it is but an illusion in the same sense in which heat is an illusion. There is nothing to heat (of a gas), except the mean kinetic energy of the molecule movements. There is nothing else there. Now, similarly, a scientific account of consciousness will be reductionistic. It will show that there is nothing to consciousness except the behavior of the neurons. There is nothing else there. And this is really the death blow to the idea that there will be a causal relation between the behavior of the micro-elements, in this case neurons, and the conscious states of the system.*

Answer to Thesis 8.

The concept of reduction is one of the most confused notions in science and philosophy. In the literature on the philosophy of science, I found at least half a dozen different concepts of reductionism. It seems to me that the notion has probably outlived its usefulness. What we want from science are general laws and causal explanations. Now, typically when we get a causal explanation, say of a disease, we can redefine the phenomenon in terms of the cause and so reduce the phenomenon to its cause. For example, instead of defining measles in terms of its symptoms, we redefine it in terms of the virus that causes the symptoms. So, measles is reduced to the presence of a certain kind of virus. There is no factual difference between saying, ''the virus causes the symptoms which constitute the disease'', and ''the presence of the virus just is the presence of the disease, and the disease causes the symptoms.'' The facts are the same in both cases. The reduction is just a matter of different terminology. This is the point: What we want to know is, what are the facts?

In the case of reduction and causal explanations of the sort that I just gave, it seems to me that there are two sorts of reductions — those that eliminate the phenomenon being reduced by showing that there is really nothing there

in addition to the features of the reducing phenomena, and those that do not eliminate the phenomenon but simply give a causal explanation of it. I don't suppose that this is a very precise distinction but some examples of it will make it intuitively clear. In the case of heat, we need to distinguish between the movement of the molecules with a certain kinetic energy on the one hand and the subjective sensations of heat on the other. There is nothing there except the molecules moving with a certain kinetic energy and this then causes in us the sensations that we call sensations of heat. The reductionist account of heat carves off the subjective sensations and defines heat as the kinetic energy of the molecule movements. We have an eliminative reduction of heat because there is no objective phenomenon there except the kinetic energy of the molecule movements. Analogous remarks can be made about color. There is nothing there but the differential scattering of light and these cause in us the experiences that we call color experiences. But there isn't any color phenomenon there beyond the causes in the form of light reflectances and their subjective effects on us. In such cases, we can do an eliminative reduction of heat and color. We can say there is nothing there but the physical causes and these cause the subjective experiences. Such reductions are eliminative reductions in the sense that they get rid of the phenomenon that is being reduced. But in this respect they differ from the reductions of solidity to the vibratory movement of molecules in lattice structures. Solidity is a causal property of the system which cannot be eliminated by the reduction of solidity to the vibratory movements of molecules in lattice type structures.

But now why can't we do an eliminative reduction of consciousness in the way that we did for heat and color? The pattern of the facts is parallel: For heat and color we have physical causes and subjective experiences. For consciousness we have physical causes in the form of brain processes and the subjective experience of consciousness. So it seems we should reduce consciousness to brain processes. And of course we could if we wanted to, at least in this trivial sense: We could redefine the word 'consciousness' to mean the neurobiological causes of our subjective experiences. But if we did, we would still have the subjective experiences left over, and the whole point of having the concept of consciousness was to have a word to name those subjective experiences. The other reductions were based on carving off the subjective experience of heat, color, etc., and redefining the notion in terms of the causes of those experiences. But where the phenomenon that we are discussing is the subjective experience itself, you cannot carve off the subjective experience and redefine the notion in terms of its causes, without losing the whole point of having the concept in the first place. The asymmetry between heat and color on the one hand and consciousness on the other has not to do with the facts in the world, but rather with our definitional practices. We need a word to refer to ontologically subjective phenomena of awareness or sentience. And we would lose that feature of the concept of consciousness if we were to redefine the word in terms of the causes of our experiences.

You can't make the appearance–reality distinction for conscious states themselves, as you can for heat and color, because for conscious states, the existence of the appearance is the reality in question. If it seems to me I am conscious then I am conscious. And that is not an epistemic point. It does not imply that we have certain knowledge of the nature of our conscious states. On the contrary we are frequently mistaken about our own conscious states, as for example in the case of phantom limb pains. It is a point about the ontology of conscious states.

When we study consciousness scientifically, I believe we should forget about our old obsession with reductionism and seek causal explanations. What we want is a causal explanation of how brain processes cause our conscious experiences. The obsession with reductionism is a hangover from an earlier phase in the development of scientific knowledge.

Thesis 9

Any genuinely scientific account of consciousness must be an information processing account. That is, we must see consciousness as consisting of a series of information processes, and the standard apparatus that we have for accounting for information processing in terms of symbol manipulation by a computing device must form the basis of any scientific account of consciousness.

Answer to Thesis 9.

I have actually, in a number of works, answered this mistake in detail (Cf. Ref. [8], see also [1,2]). But for present purposes, the essential thing to remember is this: Consciousness is an intrinsic feature of certain human and animal nervous systems. The problem with the concept of 'information processing' is that information processing is typically in the mind of an observer. For example, we treat a computer as a bearer and processor of information, but intrinsically, the computer is simply an electronic circuit. We design, build and use such circuits because we can interpret their inputs, outputs, and intermediate processes as information bearing, but in such a case the information in the computer is in the eye of the beholder, it is not intrinsic to the computational system. What goes for the concept of information goes a fortiori for the concept of 'symbol manipulation'. The electrical state transitions of a computer are symbol manipulations only relative to the attachment of a symbolic interpretation by some designer, programmer or user. The reason we cannot analyze consciousness in terms of information processing and symbol manipulation is that consciousness is intrinsic to the biology of nervous systems, information processing and symbol manipulation are observer relative.

For this reason, any system at all can be interpreted as an information processing system. The stomach processes

information about digestion, the falling body processes information about time, distance, and gravity. And so on.

The exception to the claim that information processing is observer relative are precisely cases where some conscious agent is thinking. If I as a conscious agent think, consciously or unconsciously, ''2 + 2 = 4'', then the information processing and symbol manipulation are intrinsic to my mental processes, because they are the processes of a conscious agent. But in that respect my mental processes differ from my pocket calculator adding 2 + 2 and getting 4. The addition in the calculator is not intrinsic to the circuit, the addition in me is intrinsic to my mental life.

The result of these observations is that in order to make the distinction between the cases which are intrinsically information bearing and symbol manipulating from those which are observer relative we need the notion of consciousness. Therefore, we cannot explain the notion of consciousness in terms of information processing and symbol manipulations.

Conclusion

There are other mistakes I could have discussed, but I hope the removal of these I listed will actually help us make progress in the study of consciousness. My main message is that we need to take consciousness seriously as a biological phenomenon. Conscious states are caused by neuronal processes, they are realized in neuronal systems and they are intrinsically inner, subjective states of awareness or sentience.

We want to know how they are caused by and realized in the brain. Perhaps they can also be caused by some sort of chemistry different from brains altogether, but until we know how brains do it we are not likely to be able to produce it artificially in other chemical systems. The mistakes to avoid are those of changing the subject — thinking that consciousness is a matter of information processing or behavior, for example — or not taking consciousness seriously on its own terms. Perhaps above all, we need to forget about the history of science, and get on with producing what may turn out to be a new phase in that history.

References

[1] J.R. Searle, Minds, Brains and Science, Harvard University Press, Cambridge, MA, 1984.
[2] J.R. Searle, The Rediscovery of the Mind, MIT Press, Cambridge, MA, 1992.
[3] Rene Descartes, Meditations on First Philosophy, especially Third Meditation (For example, ''But in order for a given idea to contain such and such objective reality, it must surely derive it from some cause which contains at least as much formal reality as there is objective reality in the idea.'') The Philosophical Writings of Descartes, Vol. II, translated by J. Cottingham, R. Stoothoff, D. Murdoch, Cambridge University Press, Cambridge, MA, 1984.
[4] Thomas Nagel, What Is It Like to be a Bat?, The Philosophical Review LXXXIII (4) (1974) 435–450.
[5] Francis Crick, The Astonishing Hypothesis: The Scientific Search for the Soul, Simon and Schuster, New York, 1994; Gerald Edelman, the Remembered Present: A Biological Theory of Consciousness, Basic Books, New York, 1989.
[6] J.R. Searle, The Rediscovery of the Mind, MIT Press, Cambridge, MA, 1992, Chapter 3.
[7] J.R. Searle, The Rediscovery of the Mind, MIT Press, Cambridge, 1992, Ch. 5, p. 111ff.
[8] J.R. Searle, Minds, Brains, and Programs, in: Behavioral and Brain Sciences, 1980, Vol. 3.