Consciousness

John R. Searle

Abstract

Until very recently, most neurobiologists did not regard consciousness as a suitable topic for scientific investigation. This reluctance was based on certain philosophical mistakes, primarily the mistake of supposing that the subjectivity of consciousness made it beyond the reach of an objective science. Once we see that consciousness is a biological phenomenon like any other, then it can be investigated neurobiologically. Consciousness is entirely caused by neurobiological processes and is realized in brain structures. The essential trait of consciousness that we need to explain is unified qualitative subjectivity. Consciousness thus differs from other biological phenomena in that it has a subjective or first-person ontology, but this subjective ontology does not prevent us from having an epistemically objective science of consciousness. We need to overcome the philosophical tradition that treats the mental and the physical as two distinct metaphysical realms. Two common approaches to consciousness are those that adopt the building block model, according to which any conscious field is made of its various parts, and the unified field model, according to which we should try to explain the unified character of subjective states of consciousness. These two approaches are discussed and reasons are given for preferring the unified field theory to the building block model. Some relevant research on consciousness involves the subjects of blindsight, the split-brain experiments, binocular rivalry, and gestalt switching.

I. Resistance to the Problem

As recently as two decades ago there was little interest among neuroscientists, philosophers, psychologists and cognitive scientists generally in the problem of consciousness. Reasons for the resistance to the problem varied from discipline to discipline. Philosophers had turned to the analysis of language, psychologists had become convinced that a scientific psychology must be a science of behavior, and cognitive scientists took their research program to be the discovery of the computer programs in the brain that, they thought, would explain cognition. It seemed especially puzzling that neuroscientists should be reluctant to deal with the problem of consciousness, because one of the chief functions of the brain is to cause and sustain conscious states. Studying the brain without studying consciousness would be like studying the stomach without studying digestion, or studying genetics without studying the inheritance of traits. When I first got interested in this problem seriously and tried to discuss it with brain scientists, I found that most of them were not interested in the question.

The reasons for this resistance were various but they mostly boiled down to two. First, many neuroscientists felt -- and some still do -- that consciousness is not a suitable subject for neuroscientific investigation. A legitimate brain science can study the microanatomy of the Purkinje cell, or attempt to discover new neurotransmitters, but consciousness seems too airy-fairy and touchy-feely to be a real scientific subject. Others did not exclude consciousness from scientific investigation, but they had a second reason: "We are not ready" to tackle the problem of consciousness. They may be right about that, but my guess is that a lot of people in the early 1950s thought we were not ready to tackle the problem of the molecular basis of life and heredity. They

were wrong; and I suggest for the current question, the best way to get ready to deal with a research problem may be to try to solve it.

There were, of course, famous earlier twentieth century exceptions to the general reluctance to deal with consciousness, and their work has been valuable. I am thinking in particular of the work of Sir Arthur Sherrington, Roger Sperry, and Sir John Eccles.

Whatever was the case 20 years ago, today many serious researchers are attempting to tackle the problem. Among neuroscientists who have written recent books about consciousness are Cotterill (1998), Crick (1994), Damasio (1999), Edelman (1989, 1992), Freeman (1995), Gazzaniga (1988), Greenfield (1995), Hobson (1999), Libet (1993), and Weiskrantz (1997). As far as I can tell, the race to solve the problem of consciousness is already on. My aim here is not to try to survey this literature but to characterize some of the neurobiological problems of consciousness from a philosophical point of view.

II. Consciousness as a Biological Problem

What exactly is the neurobiological problem of consciousness? The problem, in its crudest terms, is this: How exactly do brain processes cause conscious states and how exactly are those states realized in brain structures? So stated, this problem naturally breaks down into a number of smaller but still large problems: What exactly are the neurobiological correlates of conscious states (NCC), and which of those correlates are actually causally responsible for the production of consciousness? What are the principles according to which biological phenomena such as neuron firings can bring about subjective states of sentience or awareness? How do those principles relate to the already well understood principles of biology? Can we explain consciousness with the existing theoretical apparatus or do we need some revolutionary new theoretical concepts to explain it? Is consciousness localized in certain regions of the brain or is it a global phenomenon? If it is confined to certain regions, which ones? Is it correlated with specific anatomical features, such as specific types of neurons, or is it to be explained functionally with a variety of anatomical correlates? What is the right level for explaining consciousness? Is it the level of neurons and synapses, as most researchers seem to think, or do we have to go to higher functional levels such as neuronal maps (Edelman 1989, 1992), or whole clouds of neurons (Freeman 1995), or are all of these levels much too high and we have to go below the level of neurons and synapses to the level of the microtubules (Penrose 1994 and Hameroff 1998a, 1998b)? Or do we have to think much more globally in terms of Fourier transforms and holography (Pribram 1976, 1991, 1999)?

As stated, this cluster of problems sounds similar to any other such set of problems in biology or in the sciences in general. It sounds like the problem concerning microorganisms: How, exactly, do they cause disease symptoms and how are those symptoms manifested in patients? Or the problem in genetics: By what mechanisms exactly does the genetic structure of the zygote produce the phenotypical traits of the mature organism? In the end I think that is the right way to think of the problem of consciousness -- it is a biological problem like any other, because consciousness is a biological phenomenon in exactly the same sense as digestion, growth, or photosynthesis. But unlike other problems in biology, there is a persistent series of philosophical problems that surround the problem of consciousness and before addressing some current research I would like to address some of these problems.

III. Identifying the Target: The Definition of Consciousness.

One often hears it said that "consciousness" is frightfully hard to define. But if we are talking about a definition in common sense terms, sufficient to identify the target of the investigation, as opposed to a precise scientific definition of the sort that typically comes at the end of a scientific investigation, then the word does not seem to me hard to define. Here is the definition : Consciousness consists of inner, qualitative, subjective states and processes of sentience or awareness. Consciousness, so defined, begins when we wake in the morning from a dreamless sleep - and continues until we fall asleep again, die, go into a coma or otherwise become "unconscious." It includes all of the enormous variety of the awareness that we think of as characteristic of our waking life. It includes everything from feeling a pain, to perceiving objects visually, to states of anxiety and depression, to working out cross word puzzles, playing chess, trying to remember your aunt's phone number, arguing about politics, or to just wishing you were somewhere else. Dreams on this definition are a form of consciousness, though of course they are in many respects quite different from waking consciousness.

This definition is not universally accepted and the word consciousness is used in a variety of other ways. Some authors use the word only to refer to states of self consciousness*,* i.e. the consciousness that humans and some primates have of themselves as agents. Some use it to refer to the second-order mental *states about other mental states;* so according to this definition, a pain would not be a conscious state, but worrying about a pain would be a conscious state. Some use "consciousness" behavioristically to refer to any form of complex intelligent behavior. It is, of course, open to anyone to use any word anyway he likes, and we can always redefine consciousness as a technical term. Nonetheless, there is a genuine phenomenon of consciousness in the ordinary sense, however we choose to name it; and it is that phenomenon that I am trying to identify now, because I believe it is the proper target of the investigation.

Consciousness has distinctive features that we need to explain. Because I believe that some, not all, of the problems of consciousness are going to have a neurobiological solution, what follows is a shopping list of what a neurobiological account of consciousness should explain.

IV. The Essential Feature of Consciousness: The Combination of Qualitativeness, Subjectivity and Unity

Consciousness has three aspects that make it different from other biological phenomena, and indeed different from other phenomena in the natural world. These three aspects are qualitativeness, subjectivity, and unity. I used to think that for investigative purposes we could treat them as three distinct features, but because they are logically interrelated, I now think it best to treat them together, as different aspects of the same feature. They are not separate because the first implies the second, and the second implies the third. I discuss them in order.

Qualitativeness

Every conscious state has a certain qualitative feel to it, and you can see this clearly if you consider examples. The experience of tasting beer is very different from hearing Beethoven's Ninth Symphony, and both of those have a different qualitative character from smelling a rose or seeing a sunset. These examples illustrate the different qualitative features of conscious experiences. One way to put this point is to say that for every conscious experience there is something that it feels like, or something that it is like to have that conscious experience. Nagel (1974) made this point over two decades ago when he pointed out that if bats are conscious, then there is something that "it

is like" to be a bat. This distinguishes consciousness from other features of the world, because in this sense, for a nonconscious entity such as a car or a brick there is nothing that "it is like" to be that entity. Some philosophers describe this feature of consciousness with the word qualia, and they say there is a special problem of qualia. I am reluctant to adopt this usage, because it seems to imply that there are two separate problems, the problem of consciousness and the problem of qualia. But as I understand these terms, "qualia" is just a plural name for conscious states. Because "consciousness" and "qualia" are coextensive, there seems no point in introducing a special term. Some people think that qualia are characteristic only of perceptual experiences, such as seeing colors and having sensations such as pains, but that there is no qualitative character to thinking. As I understand these terms, that is wrong. Even conscious thinking has a qualitative feel to it. There is something it is like to think that two plus two equals four. There is no way to describe it except by saying that it is the character of thinking consciously “two plus two equals four". But if you believe there is no qualitative character to thinking that, then try to think the same thought in a language you do not know well. If I think in French "deux et deux fait quatre," I find that it feels quite different. Or try thinking, more painfully, “two plus two equals one hundred eighty-seven." Once again I think you will agree that these conscious thoughts have different characters. However, the point must be trivial; that is, whether or not conscious thoughts are qualia must follow from our definition of qualia. As I am using the term, thoughts definitely are qualia.

Subjectivity

Conscious states only exist when they are experienced by some human or animal subject. In that sense, they are essentially subjective.

I used to treat subjectivity and qualitativeness as distinct features, but it now seems to me that properly understood, qualitativeness implies subjectivity, because in order for there to be a qualitative feel to some event, there must be some subject that experiences the event. No subjectivity, no experience. Even if more than one subject experiences a similar phenomenon, say two people listening to the same concert, all the same, the qualitative experience can only exist as experienced by some subject or subjects. And even if the different token experiences are qualitatively identical, that is they all exemplify the same type, nonetheless each token experience can only exist if the subject of that experience has it. Because conscious states are subjective in this sense, they have what I will call a first-person ontology, as opposed to the third-person ontology of mountains and molecules, which can exist even if no living creatures exist. Subjective conscious states have a first-person ontology (“ontology” here means mode of existence) because they only exist when they are experienced by some human or animal agent. They are experienced by some "I" that has the experience, and it is in that sense that they have a first-person ontology.

Unity

All conscious experiences at any given point in an agent's life come as part of one unified conscious field. If I am sitting at my desk looking out the window, I do not just see the sky above and the brook below shrouded by the trees, and at the same time feel the pressure of my body against the chair, the shirt against my back, and the aftertaste of coffee in my mouth, rather I experience all of these as part of a single unified conscious field. This unity of any state of qualitative subjectivity has important consequences for a scientific study of consciousness. I say more about them later on. At present I just want to call attention to the fact that the unity is already implicit in subjectivity and qualitativeness for the following reason: If you try to imagine that my conscious state is broken into 17 parts, what you imagine is not a single conscious subject with 17

different conscious states but rather 17 different centers of consciousness. A conscious state, in short, is by definition unified, and the unity will follow from the subjectivity and the qualitativeness, because there is no way you could have subjectivity and qualitativeness except with that particular form of unity.

There are two areas of current research where the aspect of unity is especially important. These are first, the study of the split-brain patients by Gazzaniga, (1998) and others (Gazzaniga, Bogen, and Sperry 1962, 1963), and second, the study of the binding problem by a number of contemporary researchers. The interest of the split-brain patients is that both the anatomical and the behavioral evidence suggest that in these patients there are two centers of consciousness that after commissurotomy are communicating with each other only imperfectly. They seem to have, so to speak, two conscious minds inside one skull.

The interest of the binding problem is that it looks like this problem might give us in microcosm a way of studying the nature of consciousness, because just as the visual system binds all of the different stimulus inputs into a single unified visual percept, so the entire brain somehow unites all of the variety of our different stimulus inputs into a single unified conscious experience. Several researchers have explored the role of synchronized neuron firings in the range of 40hz to account for the capacity of different perceptual systems to bind the diverse stimuli of anatomically distinct neurons into a single perceptual experience. (Llinas 1990, Llinas and Pare 1991, Llinas and Ribary 1993, Llinas and Ribary,1992, Singer 1993, 1995, Singer and Gray, 1995,) For example in the case of vision, anatomically separate neurons specialized for such things as line, angle and color all contribute to a single, unified, conscious visual experience of an object. Crick (1994) extended the proposal for the binding problem to a general hypothesis about the NCC. He put forward a tentative hypothesis that the NCC consists of synchronized neuron firings in the general range of 40 Hz in various networks in the thalamocortical system, specifically in connections between the thalamus and layers four and six of the cortex.

This kind of instantaneous unity has to be distinguished from the organized unification of conscious sequences that we get from short term or iconic memory. For nonpathological forms of consciousness at least some memory is essential in order that the conscious sequence across time can come in an organized fashion. For example, when I speak a sentence I have to be able to remember the beginning of the sentence at the time I get to the end if I am to produce coherent speech. Whereas instantaneous unity is essential to, and is part of, the definition of consciousness, organized unity across time is essential to the healthy functioning of the conscious organism, but it is not necessary for the very existence of conscious subjectivity.

This combined feature of qualitative, unified subjectivity is the essence of consciousness and it, more than anything else, is what makes consciousness different from other phenomena studied by the natural sciences. The problem is to explain how brain processes, which are objective third person biological, chemical and electrical processes, produce subjective states of feeling and thinking. How does the brain get us over the hump, so to speak, from events in the synaptic cleft and the ion channels to conscious thoughts and feelings? If you take seriously this combined feature as the target of explanation, I believe you get a different sort of research project from what is currently the most influential. Most neurobiologists take what I will call the building block approach: Find the NCC for specific elements in the conscious field such as the experience of color, and then construct the whole field out of such building blocks. Another approach, which I will call the unified field approach, would take the research problem to be one of explaining how the brain produces a unified field of subjectivity to start with. On the unified field approach, there

are no building blocks, rather there are just modifications of the already existing field of qualitative subjectivity. I say more about this later.

Some philosophers and neuroscientists think we can never have an explanation of subjectivity: We can never explain why warm things feel warm and red things look red. To these skeptics there is a simple answer: We know it happens. We know that brain processes cause all of our inner qualitative, subjective thoughts and feelings. Because we know that it happens we ought to try to figure out how it happens. Perhaps in the end we will fail but we cannot assume the impossibility of success before we try.

Many philosophers and scientists also think that the subjectivity of conscious states makes it impossible to have a strict science of consciousness. For, they argue, if science is by definition objective, and consciousness is by definition subjective, it follows that there cannot be a science of consciousness. This argument is fallacious. It commits the fallacy of ambiguity over the terms objective and subjective. Here is the ambiguity: We need to distinguish two different senses of the objective-subjective distinction. In one sense, the epistemic sense ("epistemic" here means having to do with knowledge), science is indeed objective. Scientists seek truths that are equally accessible to any competent observer and that are independent of the feelings and attitudes of the experimenters in question. An example of an epistemically objective claim would be "Bill Clinton weighs 210 pounds". An example of an epistemically subjective claim would be "Bill Clinton is a good president". The first is objective because its truth or falsity is settleable in a way that is independent of the feelings and attitudes of the investigators. The second is subjective because it is not so settleable. But there is another sense of the objective-subjective distinction, and that is the ontological sense ("ontological" here means having to do with existence). Some entities, such as pains, tickles, and itches, have a subjective mode of existence, in the sense that they exist only as experienced by a conscious subject. Others, such as mountains, molecules and tectonic plates have an objective mode of existence, in the sense that their existence does not depend on any consciousness. The point of making this distinction is to call attention to the fact that the scientific requirement of epistemic objectivity does not preclude ontological subjectivity as a domain of investigation. There is no reason whatever why we cannot have an objective science of pain, even though pains only exist when they are felt by conscious agents. The ontological subjectivity of the feeling of pain does not preclude an epistemically objective science of pain. Though many philosophers and neuroscientists are reluctant to think of subjectivity as a proper domain of scientific investigation, in actual practice, we work on it all the time. Any neurology textbook will contain extensive discussions of the etiology and treatment of such ontologically subjective states as pains and anxieties.

V. Some Other Features

To keep this list short, I mention some other features of consciousness only briefly.

Feature 2:Intentionality

Most important, conscious states typically have "intentionality," that property of mental states by which they are directed at or about objects and states of affairs in the world. Philosophers use the word intentionality not just for "intending" in the ordinary sense but for any mental phenomena at all that have referential content. According to this usage, beliefs, hopes, intentions, fears, desires and perceptions all are intentional. So if I have a belief, I must have a belief about something. If I have a normal visual experience, it must seem to me that I am actually seeing something, etc. Not all conscious states are intentional and not all intentionality is conscious; for

example, undirected anxiety lacks intentionality, and the beliefs a man has even when he is asleep lack consciousness then and there. But I think it is obvious that many of the important evolutionary functions of consciousness are intentional: For example, an animal has conscious feelings of hunger and thirst, engages in conscious perceptual discriminations, embarks on conscious intentional actions, and consciously recognizes both friend and foe. All of these are conscious intentional phenomena and all are essential for biological survival. A general neurobiological account of consciousness will explain the intentionality of conscious states. For example, an account of color vision will naturally explain the capacity of agents to make color discriminations.

Feature 3, The Distinction Between Center and Periphery of Attention.

It is a remarkable fact that within my conscious field at any given time I can shift my *attention* at will from one aspect to another. So for example, right now I am not paying any attention to the pressure of the shoes on my feet or the feeling of the shirt on my neck. But I can shift my attention to them any time I want. There is already a fair amount of useful work done on attention.

Feature 4. All Human Conscious Experiences Are in Some Mood or Other.

There is always a certain flavor to one's conscious states, always an answer to the question "How are you feeling?". The moods do not necessarily have names. Right now I am not especially elated or annoyed, not ecstatic or depressed, not even just blah. But all the same I will become acutely aware of my mood if there is a dramatic change, if I receive some extremely good or bad news, for example. Moods are not the same as emotions, though the mood we are in will predispose us to having certain emotions.

We are, by the way, closer to having pharmacological control of moods with such drugs as Prozac than we are to having control of other internal features of consciousness.

Feature 5. All Conscious States Come to Us in the Pleasure/Unpleasure Dimension

For any total conscious experience there is always an answer to the question of whether it was pleasant, painful, unpleasant, neutral, etc. The pleasure/unpleasure feature is not the same as mood, though of course some moods are more pleasant than others.

Feature 6. Gestalt Structure.

The brain has a remarkable capacity to organize very degenerate perceptual stimuli into coherent conscious perceptual forms. I can, for example, recognize a face, or a car, on the basis of very limited stimuli. The best known examples of Gestalt structures come from the researches of the Gestalt psychologists.

Feature 7. Familiarity

There is in varying degrees a sense of familiarity that pervades our conscious experiences. Even if I see a house I have never seen before, I still recognize it as a house; it is of a form and structure that is familiar to me. Surrealist painters try to break this sense of the familiarity and ordinariness of our experiences, but even in surrealist paintings the drooping watch still looks like a watch, and the three-headed dog still looks like a dog.

One could continue this list, and I have done so in other writings (Searle 1992). The point now is to get a minimal shopping list of the features that we want a neurobiology of consciousness to explain. In order to look for a causal explanation we need to know what the effects are that need

explanation. Before examining some current research projects, we need to clear more of the ground.

VI. The Traditional Mind-Body Problem and How to Avoid It.

The confusion about objectivity and subjectivity I mentioned earlier is just the tip of the iceberg of the traditional mind-body problem. Though ideally I think scientists would be better off if they just ignored this problem, the fact is that they are as much victims of the philosophical traditions as anyone else, and many scientists, like many philosophers, are still in the grip of the traditional categories of mind and body, mental and physical, dualism and materialism, etc. This is not the place for a detailed discussion of the mind-body problem, but I need to say a few words about it so that, in the discussion that follows, we can avoid the confusions it has engendered.

The simplest form of the mind body problem is this: What exactly is the relation of consciousness to the brain? There are two parts to this problem, a philosophical part and a scientific part. I have already been assuming a simple solution to the philosophical part. The solution, I believe, is consistent with everything we know about biology and about how the world works. It is this: Consciousness and other sorts of mental phenomena are caused by neurobiological processes in the brain, and they are realized in the structure of the brain. In a word, the conscious mind is caused by brain processes and is itself a higher level feature of the brain.

The philosophical part is relatively easy but the scientific part is much harder. How, exactly, do brain processes cause consciousness and how, exactly, is consciousness realized in the brain? I want to be very clear about the philosophical part, because it is not possible to approach the scientific question intelligently if the philosophical issues are unclear. Notice two features of the philosophical solution. First, the relationship of brain mechanisms to consciousness is one of causation. Processes in the brain cause our conscious experiences. Second, this does not force us to any kind of dualism because the form of causation is bottom-up, and the resulting effect is simply a higher level feature of the brain itself, not a separate substance. Consciousness is not like some fluid squirted out by the brain. A conscious state is rather a state that the brain is in. Just as water can be in a liquid or solid state without liquidity and solidity being separate substances, so consciousness is a state that the brain is in without consciousness being a separate substance.

Notice that I stated the philosophical solution without using any of the traditional categories of "dualism," "monism," "materialism," and all the rest of it. Frankly, I think those categories are obsolete. But if we accept those categories at face value, then we get the following picture: You have a choice between dualism and materialism. According to dualism, consciousness and other mental phenomena exist in a different ontological realm altogether from the ordinary physical world of physics, chemistry, and biology. According to materialism consciousness as I have described it does not exist. Neither dualism nor materialism as traditionally construed, allows us to get an answer to our question. Dualism says that there are two kinds of phenomena in the world, the mental and the physical; materialism says that there is only one, the material. Dualism ends up with an impossible bifurcation of reality into two separate categories and thus makes it impossible to explain the relation between the mental and the physical. But materialism ends up denying the existence of any irreducible subjective qualitative states of sentience or awareness. In short, dualism makes the problem insoluble; materialism denies the existence of any phenomenon to study, and hence of any problem.

On the view that I am proposing, we should reject those categories altogether. We know enough about how the world works to know that consciousness is a biological phenomenon caused

by brain processes and realized in the structure of the brain. It is irreducible not because it is ineffable or mysterious, but because it has a first person ontology, and therefore cannot be reduced to phenomena with a third person ontology. The traditional mistake that people have made in both science and philosophy has been to suppose that if we reject dualism, as I believe we must, then we have to embrace materialism. But on the view that I am putting forward, materialism is just as confused as dualism because it denies the existence of ontologically subjective consciousness in the first place. Just to give it a name, the resulting view that denies both dualism and materialism, I call biological naturalism.

VII. How Did We Get Into This Mess? A Historical Digression

For a long time I thought scientists would be better off if they ignored the history of the mind-body problem, but I now think that unless you understand something about the history, you will always be in the grip of historical categories. I discovered this when I was debating people in artificial intelligence and found that many of them were in the grip of Descartes, a philosopher many of them had not even read.

What we now think of as the natural sciences did not really begin with Ancient Greece. The Greeks had almost everything, and in particular they had the wonderful idea of a "theory". The invention of the idea of a theory -- a systematic set of logically related propositions that attempt to explain the phenomena of some domain -- was perhaps the greatest single achievement of Greek civilization. However, they did not have the institutionalized practice of systematic observation and experiment. That came only after the Renaissance, especially in the 17th century. When you combine systematic experiment and testability with the idea of a theory, you get the possibility of science as we think of it today. But there was a feature of the seventeenth century, which was a local accident and which is still blocking our path. It is that in the seventeenth century there was a very serious conflict between science and religion, and it seemed that science was a threat to religion. Part of the way that the apparent threat posed by science to orthodox Christianity was deflected was due to Descartes and Galileo. Descartes, in particular, argued that reality divides into two kinds, the mental and the physical, *res cogitans* and *res extensa*. Descartes made a useful division of the territory: Religion had the territory of the soul, and science could have material reality. But this gave people the mistaken conception that science could only deal with objective third person phenomena, it could not deal with the inner qualitative subjective experiences that make up our conscious life. This was a perfectly harmless move in the 17th century because it kept the church authorities off the backs of the scientists. (It was only partly successful. Descartes, after all, had to leave Paris and go live in Holland where there was more tolerance, and Galileo had to make his famous recantation to the church authorities of his heliocentric theory of the planetary system.) However, this history has left us with a tradition and a tendency not to think of consciousness as an appropriate subject for the natural sciences, in the way that we think of disease, digestion, or tectonic plates as subjects of the natural sciences. I urge us to overcome this reluctance, and in order to overcome it we need to overcome the historical tradition that made it seem perfectly natural to avoid the topic of consciousness altogether in scientific investigation.

VIII. Summary Of The Argument To This Point

I am assuming that we have established the following: Consciousness is a biological phenomenon like any other. It consists of inner qualitative subjective states of perceiving, feeling and thinking. Its essential feature is unified, qualitative subjectivity. Conscious states are caused by neurobiological processes in the brain, and they are realized in the structure of the brain. To say

this is analogous to saying that digestive processes are caused by chemical processes in the stomach and the rest of the digestive tract, and that these processes are realized in the stomach and the digestive tract. Consciousness differs from other biological phenomena in that it has a subjective or first person ontology. But ontological subjectivity does not prevent us from having epistemic objectivity. We can still have an objective science of consciousness. We abandon the traditional categories of dualism and materialism, for the same reason we abandon the categories of phlogiston and vital spirits: They have no application to the real world.

IX. The Scientific Study of Consciousness

How, then, should we proceed in a scientific investigation of the phenomena involved?

Seen from the outside it looks deceptively simple. There are three steps. First, one finds the neurobiological events that are correlated with consciousness (the NCC). Second, one tests to see that the correlation is a genuine causal relation. And third, one tries to develop a theory, ideally in the form of a set of laws, that would formalize the causal relationships.

These three steps are typical of the history of science. Think, for example, of the development of the germ theory of disease. First we find correlations between brute empirical phenomena. Then we test the correlations for causality by manipulating one variable and seeing how it affects the others. Then we develop a theory of the mechanisms involved and test the theory by further experiment. For example, Semmelweis in Vienna in the 1840s found that women obstetric patients in hospitals died more often from puerperal fever than did those who stayed at home. So he looked more closely and found that women examined by medical students who had just come from the autopsy room without washing their hands had an exceptionally high rate of puerperal fever. Here was an empirical correlation. When he made these young doctors wash their hands in chlorinated lime, the mortality rate went way down. He did not yet have the germ theory of disease, but he was moving in that direction. In the study of consciousness we appear to be in the early Semmelweis phase.

At the time of this writing we are still looking for the NCC. Suppose, for example, that we found, as Francis Crick once put forward as a tentative hypothesis, that the neurobiological correlate of consciousness was a set of neuron firings between the thalamus and the cortex layers 4 and 6, in the range of 40 Hz. That would be step one. And step two would be to manipulate the phenomena in question to see if you could show a causal relation. Ideally, we need to test for whether the NCC in question is both necessary and sufficient for the existence of consciousness. To establish necessity, we find out whether a subject who has the putative NCC removed thereby loses consciousness; and to establish sufficiency, we find out whether an otherwise unconscious subject can be brought to consciousness by inducing the putative NCC. Pure cases of causal sufficiency are rare in biology, and we usually have to understand the notion of sufficient conditions against a set of background presuppositions, that is, within a specific biological context. Thus our sufficient conditions for consciousness would presumably only operate in a subject who was alive, had his brain functioning at a certain level of activity, at a certain appropriate temperature, etc. But what we are trying to establish ideally is a proof that the element is not just correlated with consciousness, but that it is both causally necessary and sufficient, other things being equal, for the presence of consciousness.

Seen from the outsider's point of view, that looks like the ideal way to proceed. Why has it not yet been done? I do not know. It turns out, for example, that it is very hard to find an exact NCC, and the current investigative tools, most notably in the form of positron emission tomagraphy

scans, CAT scans, and functional magnetic resonance imaging techniques, have not yet identified the NCC. There are interesting differences between the scans of conscious subjects and sleeping subjects with REM sleep, on the one hand, and slow wave sleeping subjects on the other. But it is not easy to tell how much of the differences are related to consciousness. Lots of things are going on in both the conscious and the unconscious subjects' brains that have nothing to do with the production of consciousness. Given that a subject is already conscious, you can get parts of his or her brain to light up by getting him or her to perform various cognitive tasks such as perception or memory. But that does not give you the difference between being conscious in general, and being totally unconscious. So, to establish this first step, we still appear to be in an early a state of the technology of brain research. In spite of all of the hype surrounding the development of imaging techniques, we still, as far as I know, have not found a way to image the NCC. With all this in mind, let us turn to some actual efforts at solving the problem of consciousness.

X. The Standard Approach to Consciousness: The Building Block Model

Most theorists tacitly adopt the building block theory of consciousness. The idea is that any conscious field is made of its various parts: the visual experience of red, the taste of coffee, the feeling of the wind coming in through the window. It seems that if we could figure out what makes even one building block conscious, we would have the key to the whole structure. If we could, for example, crack visual consciousness, that would give us the key to all the other modalities. This view is explicit in the work of Crick & Koch (1998). Their idea is that if we could find the NCC for vision, then we could explain visual consciousness, and we would then know what to look for to find the NCC for hearing, and for the other modalities, and if we put all those together, we would have the whole conscious field.

The strongest and most original statement I know of the building block theory is by Bartels & Zeki (1998, Zeki & Bartels, 1998). They see the binding activity of the brain not as one that generates a conscious experience that is unified, but rather one that brings together a whole lot of already conscious experiences . As they put it (Bartels & Zeki 1998: 2327), "[C]onsciousness is not a unitary faculty, but.. it consists of many micro-consciousnesses." Our field of consciousness is thus made up of a lot of building blocks of microconsciousnesses. “Activity at each stage or node of a processing-perceptual system has a conscious correlate. Binding cellular activity at different nodes is therefore not a process preceding or even facilitating conscious experience, but rather bringing different conscious experiences together” (Bartels & Zeki 1998: 2330).

There are at least three lines of research that are consistent with, and often used to support, the building block theory.

1. Blindsight

Blindsight is the name given by the psychologist Lawrence Weiskrantz to the phenomenon whereby certain patients with damage to V1 can report incidents occurring in their visual field even though they report no visual awareness of the stimulus. For example, in the case of DB, the earliest patient studied, if an X or an O were shown on a screen in that portion of DB's visual field where he was blind, the patient when asked what he saw, would deny that he saw anything. But if asked to guess, he would guess correctly that it was an X or an O. His guesses were right nearly all the time. Furthermore, the subjects in these experiments are usually surprised at their results. When the experimenter asked DB in an interview after one experiment, "Did you know how well you had done?", DB answered, "No, I didn't, because I couldn't see anything. I couldn't see a darn thing."

(Weiskrantz 1986: 24). This research has subsequently been carried on with a number of other patients, and blindsight is now also experimentally induced in monkeys (Stoerig and Cowey, 1997).

Some researchers suppose that we might use blindsight as the key to understanding consciousness. The argument is the following: In the case of blindsight, we have a clear difference between conscious vision and unconscious information processing. It seems that if we could discover the physiological and anatomical difference between regular sight and blindsight, we might have the key to analyzing consciousness, because we would have a clear neurological distinction between the conscious and the unconscious cases.

2. Binocular Rivalry and Gestalt Switching

One exciting proposal for finding the NCC for vision is to study cases where the external stimulus is constant but where the internal subjective experience varies. Two examples of this are the gestalt switch, where the same figure, such as the Neckar cube, is perceived in two different ways, and binocular rivalry, where different stimuli are presented to each eye but the visual experience at any instant is of one or the other stimulus, not both. In such cases the experimenter has a chance to isolate a specific NCC for the visual experience, independently of the neurological correlates of the retinal stimulus (Logothetis, 1998, Logothetis & Schall, 1989). The beauty of this research is that it seems to isolate a precise NCC for a precise conscious experience. Because the external stimulus is constant and there are (at least) two different conscious experiences A and B, it seems there must be some point in the neural pathways where one sequence of neural events causes experience A and another point where a second sequence causes experience B. Find those two points and you have found the precise NCCs for two different building blocks of the whole conscious field.

3. The Neural Correlates of Vision

Perhaps the most obvious way to look for the NCC is to track the neurobiological causes of a specific perceptual modality such as vision. In a recent article, Crick & Koch (1998) assume as a working hypothesis that only some specific types of neurons will manifest the NCC. They do not think that any of the NCC of vision are in V1 (1995). The reason for thinking that V1 does not contain the NCCs is that V1 does not connect to the frontal lobes in such a way that would make V1 contribute directly to the essential information processing aspect of visual perception. Their idea is that the function of visual consciousness is to provide visual information directly to the parts of the brain that organize voluntary motor output, including speech. Thus, because the information in V1 is recoded in subsequent visual areas and does not transmit directly to the frontal cortex, they believe that V1 does not correlate directly with visual consciousness.

XI. Doubts about the Building Block Theory

The building block theory may be right but it has some worrisome features. Most important, all the research done to identify the NCCs has been carried out with subjects who are already conscious, independently of the NCC in question. Going through the cases in order, the problem with the blindsight research as a method of discovering the NCC is that the patients in question only exhibit blindsight if they are already conscious. That is, it is only in the case of fully conscious patients that we can elicit the evidence of information processing that we get in the blindsight examples. So we cannot investigate consciousness in general by studying the difference between the blindsight patient and the normally sighted patient, because both patients are fully conscious. It might turn out that what we need in our theory of consciousness is an explanation of the conscious

field that is essential to both blindsight and normal vision or, for that matter, to any other sensory modality.

Similar remarks apply to the binocular rivalry experiments. All this research is immensely valuable but it is not clear how it will give us an understanding of the exact differences between the conscious brain and the unconscious brain, because for both experiences in binocular rivalry the brain is fully conscious.

Similarly, Crick (1996) and Crick & Koch (1998) only investigated subjects who are already conscious. What one wants to know is, how is it possible for the subject to be conscious at all? Given that a subject is conscious, his consciousness will be modified by having a visual experience, but it does not follow that the consciousness is made up of various building blocks of which the visual experience is just one.

I wish to state my doubts precisely. There are (at least) two possible hypotheses.

1. The building block theory: The conscious field is made up of small components that combine to form the field. To find the causal NCC for any component is to find an element that is causally necessary and sufficient for that conscious experience. Hence to find even one is, in an important sense, to crack the problem of consciousness.

2. The unified field theory (explained in more detail below): Conscious experiences come in unified fields. In order to have a visual experience, a subject has to be conscious already and the experience is a modification of the field. Neither blindsight, binocular rivalry nor normal vision can give us a genuine causal NCC because only already conscious subjects can have these experiences.

It is important to emphasize that both hypotheses are rival empirical hypotheses to be settled by scientific research and not by philosophical argument. Why then do I prefer hypothesis 2 to hypothesis 1? The building block theory predicts that in a totally unconscious patient, if the patient meets certain minimal physiological conditions (he is alive, the brain is functioning normally, he has the right temperature, etc.), and if you could trigger the NCC for say the experience of red, then the unconscious subject would suddenly have a conscious experience of red and nothing else. One building block is as good as another. Research may prove me wrong, but on the basis of what little I know about the brain, I do not believe that is possible. Only a brain that is already over the threshold of consciousness, that already has a conscious field, can have a visual experience of red.

Furthermore on the multistage theory of Bartels & Zeki (1998, Zeki & Bartels 1998), the microconsciousnesses are all capable of a separate and independent existence. It is not clear to me what this means. I know what it is like for me to experience my current conscious field, but who experiences all the tiny microconsciousnesses? And what would it be like for each of them to exist separately?

XII. Basal consciousness and a unified field theory

There is another way to look at matters that implies another research approach. Imagine that you wake from a dreamless sleep in a completely dark room. So far you have no coherent stream of thought and almost no perceptual stimulus. Save for the pressure of your body on the bed and the sense of the covers on top of your body, you are receiving no outside sensory stimuli. All the same there must be a difference in your brain between the state of minimal wakefulness you are now in and the state of unconsciousness you were in before. That difference is the NCC I believe we should be looking for. This state of wakefulness is basal or background consciousness.

Now you turn on the light, get up, move about, etc. What happens? Do you create new conscious states? Well, in one sense you obviously do, because previously you were not consciously aware of visual stimuli and now you are. But do the visual experiences stand to the whole field of consciousness in the part whole relation? Well, that is what nearly everybody thinks and what I used to think, but here is another way of looking at it. Think of the visual experience of the table not as an object in the conscious field the way the table is an object in the room, but think of the experience as a modification of the conscious field, as a new form that the unified field takes. As Llinas and his colleagues put it, consciousness is "modulated rather than generated by the senses" (1998:1841).

I want to avoid the part whole metaphor but I also want to avoid the proscenium metaphor. We should not think of my new experiences as new actors on the stage of consciousness but as new bumps or forms or features in the unified field of consciousness. What is the difference? The proscenium metaphor gives us a constant background stage with various actors on it. I think that is wrong. There is just the unified conscious field, nothing else, and it takes different forms.

If this is the right way to look at things (and again this is a hypothesis on my part, nothing more) then we get a different sort of research project. There is no such thing as a separate visual consciousness, so looking for the NCC for vision is barking up the wrong tree. Only the already conscious subject can have visual experiences, so the introduction of visual experiences is not an introduction of consciousness but a modification of a preexisting consciousness.

The research program that is implicit in the hypothesis of unified field consciousness is that at some point we need to investigate the general condition of the conscious brain as opposed to the condition of the unconscious brain. We will not explain the general phenomenon of unified qualitative subjectivity by looking for specific local NCCs. The important question is not what the NCC for visual consciousness is, but how does the visual system introduce visual experiences into an already unified conscious field, and how does the brain create that unified conscious field in the first place. The problem becomes more specific. What we are trying to find is which features of a system that is made up of a hundred billion discreet elements, neurons, connected by synapses can produce a conscious field of the sort that I have described. There is a perfectly ordinary sense in which consciousness is unified and holistic, but the brain is not in that way unified and holistic. So what we have to look for is some massive activity of the brain capable of producing a unified holistic conscious experience. For reasons that we now know from lesion studies, we are unlikely to find this as a global property of the brain, and we have very good reason to believe that activity in the thalamocortical system is probably the place to look for unified field consciousness. The working hypothesis would be that consciousness is in large part localized in the thalamocortical system, and that the various other systems feed information to the thalamocortical system that produces modifications corresponding to the various sensory modalities. To put it simply, I do not believe we will find visual consciousness in the visual system and auditory consciousness in the auditory system. We will find a single, unified, conscious field containing visual, auditory, and other aspects.

Notice that if this hypothesis is right, it will solve the binding problem for consciousness automatically. The production of any state of consciousness at all by the brain is the production of a unified consciousness.

We are tempted to think of our conscious field as made up of the various components - visual, tactile, auditory, the stream of thought, etc. The approach whereby we think of big things as

being made up of little things has proved so spectacularly successful in the rest of science that it is almost irresistible to us. Atomic theory, the cellular theory in biology, and the germ theory of disease are all examples. The urge to think of consciousness as likewise made of smaller building blocks is overwhelming. But I think it may be wrong for consciousness. Maybe we should think of consciousness holistically, and perhaps for consciousness we can make sense of the claim that "the whole is greater than the sum of the parts." Indeed, maybe it is wrong to think of consciousness as made up parts at all. I want to suggest that if we think of consciousness holistically, then the aspects I have mentioned so far, especially our original combination of subjectivity, qualitativeness, and unity all into one feature, will seem less mysterious. Instead of thinking of my current state of consciousness as made up of the various bits, the perception of the computer screen, the sound of the brook outside, the shadows cast by the evening sun falling on the wall -- we should think of all of these as modifications, forms that the underlying basal conscious field takes after my peripheral nerve endings have been assaulted by the various external stimuli. The research implication of this is that we should look for consciousness as a feature of the brain emerging from the activities of large masses of neurons, and which cannot be explained by the activities of individual neurons. I am, in sum, urging that we take the unified field approach seriously as an alternative to the more common building block approach.

XIII. VARIATIONS ON THE UNIFIED FIELD THEORY

The idea that one should investigate consciousness as a unified field is not new and it goes back at at least as far as Kant's doctrine of the transcendental unity of apperception (Kant, 1787). In neurobiology I have not found any contemporary authors who state a clear distinction between what I have been calling the building block theory and the unified field theory but at least two lines of contemporary research are consistent with the approach urged here, the work of Llinas and his colleagues (Llinas, 1990, Llinas et al, 1998) and that of Tononi, Edelman and Sporns (Tononi & Edelman, 1998, Tononi, Edelman & Sporns 1998, Tononi, Sporns & Edelman, 1992). On the view of Llinas and his colleagues (1998) we should not think of consciousness as produced by sensory inputs but rather as a functional state of large portions of the brain, primarily the thalamocortical system, and we should think of sensory inputs serving to modulate a preexisting consciousness rather than creating consciousness anew. On their view consciousness is an "intrinsic" state of the brain, not a response to sensory stimulus inputs. Dreams are of special interest to them, because in a dream the brain is conscious but unable to perceive the external world through sensory inputs. They believe the NCC is synchronized oscillatory activity in the thalamocartical system (1998: 1845).

Tononi and Edelman have advanced what they call the dynamic core hypothesis (1998). They are struck by the fact that consciousness has two remarkable properties, the unity mentioned earlier and the extreme differentiation or complexity within any conscious field. This suggests to them that we should not look for consciousness in a specific sort of neuronal type, but rather in the activities of large neuronal populations. They seek the NCC for the unity of consciousness in the rapid integration that is achieved through the reentry mechanisms of the thalamocortical system. The idea they have is that in order to account for the combination of integration and differentiation in any conscious field, they have to identify large clusters of neurons that function together, that fire in a synchronized fashion. Furthermore this cluster, which they call a functional cluster, should also show a great deal of differentiation within its component elements in order to account for the different elements of consciousness. They think that synchronous firing among cortical regions between the cortex and the thalamus is an indirect indicator of this functional clustering. Then once

such a functional cluster has been identified, they wish to investigate whether or not it contains different activity patterns of neuronal states within it. The combination of functional clustering together with differentiation they submit as the dynamic core hypothesis of consciousness. They believe a unified neural process of high complexity constitutes a dynamic core. They also believe the dynamic core is not spread over the brain but is primarily in the thalamocortical regions, especially those involved in perceptual categorization and containing reentry mechanisms of the sort that Edelman discussed in his earlier books (1989, 1992). In a new study, they and their colleagues (Srinivasan et al 1999) claim to find direct evidence of the role of reentry mapping in the NCC. Like the adherents of the building block theory, they seek such NCCs of consciousness as one can find in the studies of binocular rivalry.

As I understand this view, it seems to combine features of both the building block and the unified field approach.

XIV. Conclusion

In my view the most important problem in the biological sciences today is the problem of consciousness. I believe we are now at a point where we can address this problem as a biological problem like any other. For decades research has been impeded by two mistaken views: first, that consciousness is just a special sort of computer program, a special software in the hardware of the brain; and second that consciousness was just a matter of information processing. The right sort of information processing -- or on some views any sort of information processing --- would be sufficient to guarantee consciousness. I have criticized these views at length elsewhere (Searle 1980, 1992, 1997) and do not repeat these criticisms here. But it is important to remind ourselves how profoundly anti-biological these views are. On these views brains do not really matter. We just happen to be implemented in brains, but any hardware that could carry the program or process the information would do just as well. I believe, on the contrary, that understanding the nature of consciousness crucially requires understanding how brain processes cause and realize consciousness.. Perhaps when we understand how brains do that, we can build conscious artifacts using some nonbiological materials that duplicate, and not merely simulate, the causal powers that brains have. But first we need to understand how brains do it.[1]

[1] I am indebted to many people for discussion of these issues. None of them is responsible for any of my mistakes. I especially wish to thank Samuel Barondes, Dale Berger, Francis Crick, Gerald Edelman, Susan Greenfield, Jennifer Hudin, John Kihlstrom, Jessica Samuels, Dagmar Searle, Wolf Singer, Barry Smith, and Gunther Stent.