

Theory of Mind & Darwin's Legacy

John R. Searle

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

To whom correspondence should be addressed: searle@berkeley.edu

Key Words: Consciousness, Objective/Subjective, Sociobiology, Intentionality, Unconscious

Abstract:

We do not have an adequate theory of consciousness. Both dualism and materialism are mistaken because they deny consciousness is part of the physical world. False claims: 1) Behaviorism 2) Computationalism 3) Epiphenomenalism 4) The Readiness Potential 5) Subjectivity 6) Materialism. Ontological subjectivity does not preclude epistemic objectivity. Observer relative phenomena are created by consciousness, but consciousness is not itself observer relative. Consciousness consists of feeling, sentience, or awareness with: 1) qualitiveness 2) ontological subjectivity 3) unified conscious field 4) intentionality 5) intentional causation. All conscious states are caused by lower level neurobiological processes in the brain, and they are realized in the brain as higher level features. Efforts to get a detailed scientific account of how brain processes cause consciousness are disappointing. The Darwinian Revolution gave us a new form of explanation, two levels were substituted: a causal level, where we specify the mechanism by which the phenotype functions, and a functional level, where we specify the selectional advantage that the phenotype provides. Sociobiology attempted to explain general features of human society, ethics, etc. It failed. For the incest taboo it confuses inhibition with prohibition. It did not explain the moral force of the taboo. To explain the function of consciousness we cannot ask, What would be subtracted if we subtracted consciousness but left everything else the same? We cannot leave everything else the same,

because consciousness is necessary for higher functions of human and animal life. The unified conscious field gives the organism vastly increased power.

I have two objectives in this article: I want first to say something about our current understanding of the nature of the mind, particularly consciousness; and secondly I want to relate my account of the mind to Darwinian evolution. Many of the ideas in this article have appeared in preliminary form by me in various books and articles, particularly significant are (1-3).

I. The “Scandal” of Consciousness

Consciousness is something of a scandal because we do not have an adequate neurobiological theory of consciousness, and there are a rather large number of false claims made about it. Here, for a start, are half a dozen false claims made about consciousness in my intellectual lifetime. Several of these have been widespread and extremely influential.

1. Behaviorism (4). Consciousness as traditionally construed does not really exist, but rather there are just human and animal behaviors and dispositions to behavior.

2. Computationalism (5) (*Strong Artificial Intelligence*). Consciousness as such does not really exist, but is rather a program or a set of computer programs running in the brain.

3. Epiphenomenalism (6). Consciousness does exist, but it cannot have any real effect on the world because it is a non-physical phenomenon and as such cannot affect the physical the world.

It must be an epiphenomenon because the physical world is “causally closed.” From an evolutionary point of view it has no function.

4. *The Readiness Potential* (7). Consciousness does exist, but it has very little importance, because research on the readiness potential in the supplementary motor cortex shows that our actions are initiated prior to our becoming consciously aware of what we are doing. The brain decides to perform an action before the conscious mind can be aware of it.

5. *Objectivity and Subjectivity*. Consciousness is not a suitable subject for serious scientific investigation; it is better left to theologians and philosophers. The reason is that science is by definition objective, and consciousness is by definition subjective, therefore there cannot be a science of consciousness. This is part of the oral tradition. When I was first interested in the neurobiology of consciousness, I discussed the issue with several neurobiologists, both in Europe and the United States. Several assured me that consciousness is not a suitable subject for serious scientific investigation because of its subjectivity.

6. *Materialism* (8). If consciousness is real, it must really be something else because the final inventory of the basic ontology of the world – an inventory that includes subatomic particles, gravity, electromagnetism, the weak and strong nuclear forces, and other fundamental features of reality – is entirely material and so will not include consciousness.

I will argue that consciousness is a real biological phenomenon in the same sense as photosynthesis or digestion, and that all six of these views are more or less demonstrably false. But that raises an interesting question, Why is there so much confusion about consciousness? Why is it so hard for people to accept a naturalistic conception? We are the victims of two traditions that appear to be inconsistent with each other, but in fact they trade off on each other. One is the tradition of God, the Soul and Immortality that says consciousness is not a part of the natural world. Consciousness is not even a property of the body or the brain. On the Cartesian view it is a property of the soul and the soul is definitely not a part of the natural world. This is

why, incidentally, Descartes (9) was forced to say that animals are not conscious, because all the animal has got is a body. For Descartes, our bodies are not conscious, our brains are not conscious, only the soul is conscious. Now you might think that tradition is entirely dead, but it is not. Artificial Intelligence in its Strong AI version is an expression of the idea that the mind is not a part of the natural world. As one of the adherents of the Strong AI view wrote: the mind is something formal and abstract (10). The second tradition is usually mistakenly described as “Materialism” and often its adherents simply deny that consciousness as qualitative, subjective states really exists. An extreme statement was given by John B. Watson, one of the founders of Behaviorism

...the time has come for psychology to discard all reference to consciousness...it is neither a definable nor a usable concept, it is merely another word for the ‘soul’ of more ancient times...(11).

Both traditions share a resistance to treating real consciousness -- qualitative, subjective consciousness -- as a real biological feature of the world.

II. Objectivity, Subjectivity, and Observer Relativity

Before refuting these views and giving an account of consciousness, I need to make some distinctions. First, the distinction between objectivity and subjectivity looms very large in our intellectual culture, but unfortunately it is systematically ambiguous between an epistemic sense and an ontological sense (“epistemic” means having to do with knowledge and “ontological” means having to do with existence).

Epistemically, the distinction is between types of claims, those that can be settled as a matter of “objective” fact and those that cannot be so settled. So if I say Van Gogh died in

France, that is epistemically objective. You can settle that as a matter of “objective fact.” If I say Van Gogh was a better painter than Gauguin, that, as people like to say, is a matter of “subjective opinion.” It cannot be settled as a matter of “objective fact.” But in addition to that distinction, there is an underlying ontological distinction between modes of existence. Some entities have a mode of existence that is independent of anybody’s attitudes or consciousness. Mountains, molecules, and tectonic plates are in that sense ontologically objective. But some entities depend for their very existence on being experienced by a human or animal subject. In this sense pains, tickles and itches are ontologically subjective. Why is this distinction important? It is of crucial importance because the ontological subjectivity of a domain does not preclude epistemic objectivity of a science of that domain. You can have a completely adequate science of consciousness or mental life in general that is epistemically objective even though the entire domain is ontologically subjective.

Related to the distinction between objectivity and subjectivity is the distinction between those features of the world whose existence depends on human attitudes and those features that exist independently of anyone's attitudes. I call the former “observer relative” and the latter “observer independent” or “absolute.” Observer relative phenomena include money, property, marriage, nation states, universities and summer vacations. Observer independent phenomena include mountains, molecules, galaxies and tectonic plates. In general the natural sciences deal with observer independent phenomena; the social sciences deal with observer relative phenomena. The observer relativity of a phenomenon introduces an element of ontological subjectivity into its very existence. So the existence of money and language, for example, is observer relative and consequently contains an element of ontological subjectivity. But, and this is a crucial point, as we have already seen the ontological subjectivity of a domain does not

preclude an epistemically objective science of that domain. There can be an epistemically objective sciences of economics and linguistics, though the objects of study of the disciplines are for the most part observer relative and thus at least in part ontologically subjective. *All observer relative phenomena are created by consciousness, but the consciousness that creates them is absolute and not observer relative.* I will say something about the role of the unconscious later.

These distinctions already will enable us to answer several of the mistaken theories about consciousness. But before I do that I want to address the analysis of consciousness directly.

III. The Analysis of Consciousness

What is consciousness? What are its features? Where do we stand in its scientific investigation? It is often said consciousness is hard to define. I do not think it is hard to define. We need to make a distinction between the scientific definition that comes at the end of an investigation and the common sense definition that enables us to identify the target of the investigation. Think of water: the common sense definition is that it is clear, colorless and tasteless liquid that flows in rivers and exists in lakes and falls out of the sky in the form of rain. The scientific definition comes at the end of the investigation: it is H₂O. With consciousness we are still in the clear, colorless, tasteless liquid stage and that is sort of definition I am going to give you. Consciousness consists of all of our states of feeling or sentience or awareness. It begins when we wake from a dreamless sleep and continuous until we go to sleep again or die or otherwise become unconscious. On this definition dreams are a form of consciousness.

This then is the definition of consciousness. What are some of its features? There are many features, but for the sake of brevity I will confine myself to the five most important:

1. *Qualitativeness*. For every conscious state there is a qualitative feel to that state, something it feels like to be in that state. Think of the difference between drinking wine, listening to music and working on your income tax. Qualitativeness is not just a feature of bodily sensations and emotions but characterizes thought as well. You can see that the thought that $2+2 = 4$ has a certain characteristic feel if you try thinking the same thought in French or German. To me at least it just feels different.

2. *Ontological subjectivity*. Because of this qualitative character every conscious experience is ontologically subjective. It only exists in so far as it is experienced by human or animal subject. This feature makes consciousness seem intractable; it makes it seem a difficult subject for scientific study.

3. *Unity*. A third feature of consciousness is that there is a remarkable unity to our conscious states. In giving a lecture, I do not just hear the sound of my voice and see the people in the audience and feel a slight headache from the wine I drank the night; but I have all of those as part of a single unified conscious field. When we get to the topic of evolution we will see the centrality of unity in enabling consciousness to perform its biological functions. You never just have a single conscious experience such as the taste of the wine or the smell of the rose, but each such experience occurs as part of a total large conscious experience, which consists of the entire ontologically subjective field of consciousness occurring at that moment.

I used to think that qualitativeness, subjectivity and unity were separate features but I now think they are just different aspects of the same feature and it is the essence of consciousness. You cannot have qualitativeness without subjectivity and you cannot have subjectivity without unity. This incidentally is why the split brain experiments are so interesting

to us is because it looks like the unity of consciousness is partly disrupted if you cut the corpus callosum. So those are all three aspects of one feature and it is essence of consciousness.

There are two other features I want to mention.

4. *Intentionality*. Intentionality is that feature of the mind by which it is directed at objects and states of affairs, (intending in the ordinary sense is just one type of intentionality, along with perception, belief, desire, fear, the emotions, etc.) Not all conscious states are intentional. Think of undirected anxiety for example. But most of them are, and this feature is really what gives importance to the previous features, because intentionality is the way that the conscious organism has of relating to the environment.

5. *Intentional Causation*. The conscious mind functions causally in producing behavior, and in conscious perception the environment is causally presented to the consciousness of the agent. For example, in the presence of food, light reflected off the object will cause a conscious visual experience in the animal, and this experience will consciously motivate the animal to move toward the food.

IV. The Solution to the “Mind-Body Problem”

Suppose that I am right so far and that consciousness consists of unified, qualitative subjectivity and that it enables us to cope with the environment by way of intentionality and intentional causation. How does such a phenomenon fit in with what we know about the rest of the universe? This, you will no doubt recognize, is the traditional mind-body problem. It is supposed to be frightfully difficult, maybe impossible, to solve. I think at the level of the general relationships between the mind and the brain it has a rather simple solution. The neurological solution turns out to be very difficult and complicated, but at the level of describing the general

relationships between consciousness and neurobiology the solution is not complex. Here it is: all of our conscious states without exception are *caused by* neurobiological processes in the brain and they are all *realized in* the brain as higher level or system features. Consciousness is a feature of the brain in a way, for example, that the liquidity of the water is a feature of the system of H₂O molecules. It is not something squirted out by the H₂O molecules; it is the condition that the system of molecules is in. In the same way, my conscious state is the condition that my brain is now in. And just as a body of water can go from a liquid state to a solid state, depending on the behavior of the microelements, the H₂O molecules, so the brain can go from a conscious state to an unconscious state depending on the behavior of the microelements, the neurons.

If that is how it works in general, then why do we not get busy and figure out exactly how it works? Why is progress so slow? I do not know. I was asked to write an article for the *Annual Review of Neuroscience* (11) in which I surveyed a lot of the literature investigating the causation and the realization of consciousness in the brain. And it occurred to me that part of the difficulty is that the research techniques are at present inadequate to address the problems that really bother us. They tend to be either imaging techniques-- fMRI being the most influential, but other forms of scanning are important as well-- or they tend to be single cell recordings. And much of the most exciting work, for example on binocular rivalry, tends to be focused on very local neurobiological structures; whereas it seems me that consciousness is more likely to be a global phenomenon. Our present research techniques make it extremely difficult to investigate how consciousness is caused in the brain as a large scale phenomenon.

I distinguish between what I call the Building Block Approach, where we think of the total conscious field as made up of a set of separate conscious building blocks such as the specific types of perception, and what I call the Unified Field Approach, where we think of

specific stimuli such as perception as modifying a pre-existing conscious field. There was a period when it seemed the crucial research task was to discover the neuronal correlate of consciousness, the NCC. And indeed in this period a lot of neuronal correlates of specific forms of consciousness were discovered. But this did not solve the problem of consciousness in a way that some people had hoped. Why not? The research on blind sight or binocular rivalry or gestalt switching was always on subjects who were already consciousness. But it did not answer the question, How does the subject's brain create an entire conscious field to begin with. Two rival pictures of this research: one is that perception creates consciousness; the second, to which I adhere, is that we should think of perception as modifying the pre-existing conscious field. And the crucial question is, How does the brain create the conscious field?

Think of it this way. Imagine that you awake in a completely dark room in an unfamiliar environment. You can be totally awake and alert, though you have zero or minimal perceptual stimuli. If you get up and move about, are you creating consciousness? In a sense you are because you have conscious states you did not have before. But I think it is best to think of these experiences as modifying the pre-existing conscious field that came into existence when you became fully awake. Most of the research I am familiar with relies either on imaging techniques or single cell recordings. The difficulty is that neither of these seems adequate to get at global properties of the brain, and it may well be the case that the creation of the conscious field requires more powerful techniques than these.

V. The Refutation of the Mistaken Views

I promised at least a brief refutation of the six mistaken theories of consciousness and we now have enough material to do that.

1. *Behaviorism*. It should be an embarrassment to us that behaviorism was so influential for so long, because it is obviously false. It denies the subjectivity of consciousness. Each of us knows from his or her or her own experience that our pain is one thing and the observable pain behavior is something else.

2. *Computationalism*. We know that the implemented computer program is not by itself sufficient for mental processes whether conscious or unconscious, because the program is defined purely syntactically in terms of symbol manipulation whereas mental processes have actual content. Syntax by itself is not sufficient for semantic content. I demonstrated this a generation ago with the so-called Chinese Room thought experiment (13). Imagine that you are carrying out the steps in a program for answering questions in Chinese or some other language you do not understand. You might give the right answers, but all the same you do not understand Chinese. Carrying out the computational steps is not sufficient for understanding.

I think the argument is conclusive but the material in this paper gives us a much deeper argument. Except for computations carried out by a conscious agent *computation is observer relative*. You cannot explain consciousness as computation, because a process is computational only relative to some conscious agent. Either a conscious agent is carrying out a computation, such as adding 2+2 to get 4, or a conscious agent is using or can use a piece of machinery such as a calculator where he can *interpret* the results as arithmetical. Such computations are always observer relative. And remember observer relativity does not imply epistemic subjectivity. It is an epistemically objective fact that I am writing this using the Word Program, and the program is implemented electronically, but "Word Program" does not name an electrical phenomenon.

3. *Epiphenomenalism*. We have literally thousands of years of experiences of human and animal consciousness causing behavior. The problem is to explain how it could, given its subjective

ontology. Here is how it works. Consider a simple act like raising my arm. My intention-in-action causes my arm to go up. But we know independently that anything that causes my arm to go up in that way must cause the secretion of acetylcholine at the axon endplates of the motor neurons. No acetylcholine no arm going up. But that means that the conscious intention-in-action has to be a biochemical phenomenon. There is no way it is going to produce the secretion of acetylcholine unless it is itself realized in a biological structure. One in the same event, my conscious intention-in-action has a level of description where it is qualitative, subjective, and part of a unified subjective conscious field, and another level of description where it is a neurobiological process realized in the brain.

Part of our difficulty in understanding this is that we are stuck with the traditional vocabulary that contains the traditional mistakes, the vocabulary of the mind and body, and dualism and materialism. What I am trying to convey with this very simple example is even for very simple conscious activities like raising your arm the traditional categories are obsolete, because you have to have the concept of a single event which has both subjective, qualitative, mentalistic features and biochemical features. And this is familiar in nature that you have the same phenomenon, the same system, with different levels of description. My car engine has a level of description where explosions occur in the cylinder that drive the piston and another level of description where individual hydrocarbon molecules oxidize. We find it difficult to appreciate these levels where the mind is concerned because one of those levels of description has such a sordid history. The dualistic tradition has given the mental level of description a bad name because it makes it appear that our mental life is not part of our ordinary biological existence.

4. The Readiness Potential. In these experiments subjects were asked to perform a trivial act such as pushing a button and to observe on a clock exactly when they undertook to do it. Some

200 to 300 ms before they were aware they had decided to do it there was an increased activity in the supplementary motor area. Incredible claims were made for these data, such as for example that they disproved free will and showed that our brain decides to perform actions before our conscious mind does (14). Recent experiments show that the original experiments were flawed. If you ask the subjects to look at a clock and decide *not* to perform an action you get the same readiness potential. As far as we can tell the readiness potential was produced by watching the clock. Take away the clock and there is no readiness potential (15).

I believe the history of the readiness potential is an unfortunate chapter in recent scientific history and it raises the question, Why were people so eager to believe these implausible conclusions? The answer I think is that they wanted to discredit consciousness. Consciousness has typically been an embarrassment to the natural sciences, and in these cases it looks like we have scientific proof that consciousness does not really matter very much for our behavior.

5. *Objective/Subjective Argument*. This argument is an obvious fallacy of ambiguity over the two senses of objective and subjective that I have explained. Science is indeed epistemically objective. But there is nothing in epistemic objectivity that prevents the investigation of a domain that is ontologically subjective.

6. *Materialism*. These categories of materialism, mentalism, dualism etc. have all become obsolete. Of course ultimate reality is as described by the natural sciences and thus is “material.” There is nothing in this concept of the material that prevents subjective, qualitative consciousness from being as much a biological phenomenon as digestion, mitosis, or photosynthesis.

VII. The Darwinian Revolution

A remarkable thing about the development of knowledge is that we get not just new explanations but new *forms* of explanation. And, to me, one of the most fascinating things about Darwinian Revolution is that we got a form of explanation that previously was unknown or certainly unappreciated. The idea was that in addition to the level of explanation of traditional Aristotelian biology, where you had a teleological explanation of a phenotype, we substituted for that explanation two different levels. Aristotle thought there were such things as Final Causes, teleological causes, where the explanation is given by specifying the goal, aim, or telos of the phenomenon to be explained. So if you want to explain why fish have the shape they do, why are fish not shaped the way a brick is shaped, or why plants turn their leaves toward the sun, you point out that the purpose of all of this is to enable the fish to swim better or enable the plant to survive. And it is this teleological goal that provides the explanation. The Darwinian Revolution produced a substitution of two different levels of explanation. Instead of saying the plant turned its leaves toward the sun because it has the goal of survival, we substitute two levels of explanation, a causal or mechanical explanation and a functional explanation. At the mechanical level, the plant has variable secretions of the growth hormone auxin and these variable secretions of auxin turn the leaves toward the sun. And at the second functional level, plants that turn their leaves towards the sun are more likely to survive than plants that do not. Notice that survival still functions in the explanation, but survival is no longer the goal that the plant has; it is just something that happens. So we have inverted the conditional. Instead of saying in order to survive the plant has to turn its leaves towards the sun, we now say the plant will turn its leaves toward the sun because of the chemical secretions; and because it turns its leaves toward the sun, it is more likely to survive than if it did not. For the traditional

Aristotelian final cause, or teleological cause, you substitute two levels. Survival still functions but it no longer functions as the goal which explains the phenomenon; it is just something that happens. This introduces another element to the explanation: the diachronic element. This kind of explanation only works over periods of time.

VII. The Failure of Sociobiology

There are limits with what you can do with Darwinian modes of explanation. Some decades ago there was a movement called Sociobiology which was going to take this model of explanation and explain human culture, morality, civilization, philosophy and pretty much everything else. The inventor of this was E.O. Wilson (16); I had a chance to debate him and several other proponents, sometimes in print, but more often in conferences. The failure of Sociobiology is revealing to us. I have just endorsed the Darwinian mode of explanation, but what are its limits? Why did Sociobiology fail? It may sound question begging to assume it failed, so I want to explain a little bit what its limitations were.

The crudest limitation that sociobiological methods had was that they were trying to explain specific features of human culture and society, hence the name Sociobiology, but the mode of explanation had to be consistent with the fact that there has been no major change in the human gene pool over the past 30,000 years. That figure incidentally comes from physical anthropologists in Berkeley, especially Sherwood Washburn (maybe the figures are inaccurate I will discuss that in a moment), but now if we are going to explain human societies think of the enormous variations in human society over the past 30,000 years. So if we want to explain things like the rise of Fascism or the democratic society produced by the Enlightenment or the existence of the Enlightenment itself, then it looks like we have too crude an analytic tool to

work with. Suppose the figure is wrong, suppose it is wrong by a factor of ten, suppose the human gene pool is what it is over the past three thousand years, the same problem still arises. There is still too much variety over the past three thousand years for a single mode of explanation to explain it all.

It might seem the answer is to concentrate on cultural universals, and that is what Wilson did. For example, all societies have an incest taboo, and Wilson thought Sociobiology provided an explanation for the taboo. Cultural anthropologists were somewhat outraged at this and they pointed out, among other things, that the incest taboo takes many different forms in different cultures. In some cultures it is forbidden to marry your cousin and in other cultures it is not. But the answer I think is that that the incest taboo always forbids brother-sister, father- daughter, mother-son sexual relations; and that is the core of the taboo. I debated Wilson once at the University of Michigan (17), and he claimed that Sociobiology has shown why incest is evil: for the first time we have an explanation of a universal feature of human morality, we have an explanation of why incest is evil. How is the explanation supposed to work on the analogy with the explanation I gave you of the two level explanations for why plants turn their leaves toward the sun? Studies have shown that children brought up in close proximity to people of the opposite sex lack sexual desire for those. The favorite example comes from Israeli Kibbutzim. It turns out that the children brought up in one Kibbutz tend to lack sexual desire for other people brought up in that same Kibbutz. They are brought up in very close proximity, especially in the communal nursery. In general, they tend to have greater sexual desires for people brought up in a different Kibbutz on the other side of the hill. And Wilson said that what happens is that close proximity during the period of development leads to a type of aversion; this sexual aversion gives the mechanical or causal explanation of the incest taboo. What is the functional

explanation? And to answer that we have to ask, Why do we have bi-parental reproduction in the first place? The answer is that you get a much better genetic result if you mix the genes. If we just reproduced like the amoeba by fission or with incest, we would not get the advantages of mixing the genes. So you have the functional level of explanation: the advantages of bi-parental reproduction and many of those advantages are lost if you reproduce by incest. And you have the causal level of explanation which is the aversion.

This seems to me an inadequate explanation of a universal prohibition. To begin with, *inhibition* does not explain *prohibition*. If it were right that humans had an innate inhibition for sexual congress with people whom they have been brought up in close proximity, then there would be little need for an incest taboo. There is, for example, no need for a prohibition against eating a lot of raw mud, because we are just not inclined to do it. But the absence of an inclination does not explain the power of prohibition. And the functional level does not give the specific intentional content to the prohibition. You do not get from “mixing the genes is useful” to “incest is evil.” Furthermore, when Wilson said we have at last explained why incest is evil, it seems no such explanation was in fact given. At most it would be an explanation of why it is a bad idea to have incestuous relations without taking precautions to prevent fertilization. But the idea that the moral component in the incest taboo is explained by there being a functional advantage to bi-parental reproduction and a natural inhibition that occurs when people grow up in proximity seems to me much too weak. This is an illustration of where in the end sociobiological modes of explanation failed as an account of the general structure of human society or the structure of morality. That is not to say that we cannot do more than we have done with Darwinian modes of explanation, but this particular intellectual movement seemed to me doomed from the start.

VIII. The Evolutionary Function of Consciousness

But then that raises the question that I often hear when I talk about consciousness, What is the evolutionary function of consciousness? Maybe it does not have a function. There are two classes of people I have argued with about this. One class says: consciousness must be epiphenomenal because the physical world is causally closed, and on your own account consciousness is irreducibly subjective, and therefore it is not reducible to the objective world. I have already answered this claim. A second form of argument is more revealing. Consciousness does not have a function because we can easily imagine all of human and animal life going on as it does, only minus consciousness. We can easily imagine that we all existed as unconscious zombies but performed exactly the same actions that we now perform. This possibility shows that consciousness has no essential evolutionary function. (One frequently hears this objection in conferences on consciousness. I most recently was presented with this objection by several critics at the Turing Consciousness 2012 conference at the Université du Québec à Montréal.) This is a deeper mistake and I will answer it.

We tend to hear the question, “What is the evolutionary function of consciousness? What selectional advantage does it convey?” On analogy with such question as “What is the selectional advantage of the vestibular ocular reflex? What selectional advantage does it convey?” In the case of the VOR the answer is simple: by stabilizing the retinal image it improves visual perception. We arrive at this answer by mentally subtracting the VOR and seeing what difference it would make. The problem with trying to do this with consciousness is, if you subtract consciousness you subtract roughly speaking all of our life, except such basic unconscious processes as breathing. What would be lost if we all suddenly went into a coma or

had always been in a coma? Roughly speaking, everything. The reason that many people fail to see this is that they typically hear the question, “What would be lost if we subtracted consciousness?” as having the form “What would be lost if we subtracted consciousness but left everything else exactly the same?” But of course everything else would not be the same. In fact, pretty much nothing would be the same. I could not get up from the bed in the morning, eat my food, make my way around the surface of the earth, and deal with other people. It is even a mistake to think of consciousness having “a function.” The question is like: what is the function of life? Everything in our life depends on being alive and just about everything in our lives depends on being conscious. The reason people tend to think that we might mentally subtract consciousness and leave everything else the same is that they are victims of residual dualism. They assume consciousness is not a part of the ordinary “physical” world, so we might just lift it off and leave everything else the same. But that is not how nature works. For example, it does not show that birds’ wings do not have a function, to argue that we can imagine a world in which birds are powered by rocket engines. The fact that we can imagine zombies that behaved like us without consciousness is like imagining that birds fly with rocket engines. It is irrelevant to giving an evolutionary explanation. The way that nature works for beings like us is that we have enormous power added by the existence of consciousness, by the existence of qualitative subjective conscious states.

Our typical vocabulary makes it difficult to state or appreciate this point because we tend to think in terms like “information” and “representation,” and it is tempting to say things like “We have much more efficient information processing and more effective forms of representations than we would if we were unconscious.” This does not do justice to the power of consciousness. Our conscious perceptions and actions do not just give us “representations.”

They give us direct *presentations* of the environment. Thanks to consciousness, I have an immediate awareness of the environment around me. And thanks to the unified consciousness field, I have a prodigious amount of presentational content in the conscious field. If you look around you for a moment and think how many things you can be consciously aware of right now, you would never know when to stop. And remember you have not only the immediate presentational character, but you also have an immediate sense of your most recent past and your intended future. You are constantly using the unified conscious field to change yourself and the environment.

I still have not adequately conveyed the enormous powers that consciousness gives us. The organism can coordinate an enormous number of simultaneous inputs within a single conscious field and can coordinate its behavior both in light of the sensory inputs and in light of its goals and the possible means for achieving its goals. It coordinates all of this in a way that simultaneously represents the past in the form of short term and long-term memory, and anticipates the future in the form of its plans, goals, and intentions. In the case of humans, consciousness also gives us the capacity for cooperating with con-specifics; and the specific form of human cooperation takes is the conscious use of language. The use of language gives us the distinguishing features of human civilization: money, property, government, social organizations, etc. -- all of these are the results of conscious application of linguistic representations.

IX. The Unconscious

What should we say about the unconscious? It is essential to distinguish *unconscious* mental phenomena from *non-conscious* neurobiological processes that make consciousness

possible. For example, even when I am sound asleep I have an *unconscious* belief that George Washington was the first president. When I see anything, there are *non-conscious* feedback mechanisms between V1 and the Lateral Geniculate Nucleus. These are essential to perception, but they have no psychological reality at all. What is the difference? The reason that we can say that my unconscious belief is psychologically real is that I am able to bring them to consciousness. So the notion of an unconscious mental state is parasitic on the notion of the conscious, because an unconscious belief, if it genuinely has psychological reality, has to be the kind of thing that you could in principle bring to consciousness. I have to say “in principle” because access to consciousness may be blocked by brain damage or repression or forgetfulness or other things, but it cannot be something that is not even in principle accessible to consciousness, because if so it has no psychological reality.

It is important to emphasize this because in the early days of Cognitive Science the explanatory paradigm consisted in large part in attempts to get computational theories of cognitive capacities (18). We were told that the computational level has psychological reality, but it is at an intermediate level between that of common sense psychology, conscious and otherwise, and the level of neurobiology. I want to say that there is no such level. You cannot make any sense of the idea that there is a psychological reality at that level if it does not have those features that constitute accessibility to consciousness in principle.

I think the paradigm is changing in Cognitive Science. I think we are moving away from a computational Cognitive Science to cognitive neuroscience. I welcome that move. I think the future of this entire research project lies in a better understanding of the brain.

References

1. Searle JR (1996) *Minds, Brains and Science*. Cambridge, MA, Harvard University Press.
2. Searle JR (1992) *The Rediscovery of the Mind*. Cambridge, MA, A Bradford Book.
3. Searle JR (2005) *Mind: A Brief Introduction*. Oxford, Oxford University Press.
4. Skinner BF (1992) *Verbal Behavior*. Acton, MA, Copley Publishing Group.
5. Minsky M, Papert S (1987) *Society of Mind*. New York, NY, Simon and Schuster.
6. Chalmers D (1996) *The Conscious Mind: In Search of a Fundamental Theory*. Cambridge, MA, Oxford University Press.
7. Libet B, Gleason CA, Wright EW, Pearl DK (1983) Time of Conscious Intention to Act in Relation to Onset of Cerebral Activity (Readiness-Potential) The Unconscious Initiation of a Freely Voluntary Act. *Brain* 106(3):623-642.
8. Dennett CD (1991) *Consciousness Explained*. Boston, MA: Little Brown and Company
9. Descartes R (1996) *Meditations on First Philosophy*. Trans. Cottingham J. Cambridge, UK, Cambridge University Press.
10. Dennett CD, Hofstadter RD (1981) *The Mind's I: fantasies and reflections on the self and soul*. New York, NY, Bantam House.
11. Watson, J.B. (1925) *Behaviorism*. New York: Norton. quoted in Baars, Bernard J. The Double Life of B.F. Skinner *Journal of Consciousness Studies* 10(1):7
12. Searle JR (2000) Consciousness *The Annual Review of Neuroscience* 23:557-578.
13. Searle JR (1980) Minds, Brain and Programs. *Behavioral and Brain Sciences* 3(3):417-457.
14. Koch C (2012) *Consciousness: Confessions of a Romantic Reductionist*. Cambridge, MA, MIT Press.

15. Trevena J, Miller J (2010) Brain preparation before a voluntary action: Evidence against unconscious movement initiation. *Consciousness and Cognition* 19(1):447-456.
16. Wilson EO (1975) *Sociobiology: The New Synthesis*. Cambridge, MA, Belknap Press of Harvard University Press.
17. Wilson. E.O. (1978-1979) *Comparative Social Theory. Tanner Lectures on Human Values* at the University of Michigan.
18. Marr D, Poggio AT, Ullman S (2010) *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. Cambridge, MA, MIT Press.