Honorary Degree Lecture, Palermo 2010

Consciousness and The Problem of Free Will

By John R. Searle

I: The Problem of Free Will

Consciousness definitely exists, but we are not sure if free will exists. That is why I prefer to say "the problem of free will," and not just "free will" in the title of this paper. There are very close connections between consciousness and the problem of free will. In order to explain those connections, I have to give at least a brief definition of each term. Consciousness is often said to be extremely hard or difficult to define, but if we are talking about a commonsense definition that just identifies the target of our investigation and not a scientific definition given in terms of the most basic neurobiological processes of consciousness, then it seems to me rather easy to give a commonsense definition. Consciousness consists of those states and events of feeling or sentience or awareness that typically begin when we awake from a dreamless sleep and continue throughout the day until we go to sleep or otherwise become unconscious again. Dreams, on this definition, are a form of consciousness. Some of the essential features of consciousness are that it is *qualitative* in that there is always a certain qualitative character to any conscious state. It is *subjective* in the sense that it only exists as experienced by a human or animal subject, and in the normal non-pathological cases, it is *unified*, it has a kind of unity in that all of my conscious experiences are experienced as part of a single conscious field.

It is much more difficult to give a neutral, non-tendentious definition of "free will", but in the sense of "free will" that I will be interested in, the problem of free will is this: are there human actions that are not determined by antecedently sufficient causal conditions? To make that more explicit, the question we are asking is this: Given that there are actions that are not random, that are done with conscious intent, and that follow from conscious decision-making, is it the case that any of these actions were not preceded by causal conditions sufficient to determine the actions. If the answer to that question is yes, then there are free actions. But it would be an odd result if there were such actions, because these would be actions which, though they admit of causal explanations given in

terms of the agent's reasons, motives, beliefs, desires, etc., all the same, the actions were not determined in the sense that the totality of the prior states of the agent, including all the neurobiological states, were not sufficient to determine that that particular action had to occur. The agent might have done something different from what he did in fact do.

The connection between the consciousness and the problem of free will is this: only for a conscious agent is there a problem of free will, and if free will does exist, it can only exist in conscious agents. I will explain both of these points. Why do we have a problem of freedom of the will at all? We do not, for example, have a problem of the freedom of digestion or freedom of perception. The reason is that there is a certain peculiar character to some of our conscious experiences. When I make up my mind to do something, when I make a decision, I typically do not experience the making of the decision as *forced* by the considerations for or against the decision. For example, a few months ago I made a decision as to which presidential candidate to vote for, but I did not experience my decision as something that I simply could not help, that there was no way that I could have decided otherwise. There was in short, an experienced conscious gap between the reasons for the action and the decision to perform the action based on the reasons. We do not always experience this gap. Sometimes we feel ourselves in the grip of a compulsion, an obsession, or an addiction. But in most normal nonpathological cases, we experience a gap between reasons and decision. Furthermore, when it comes to actually doing the thing we decided to do, we experience a second gap, or rather a continuation of the first gap, in that even though we have already decided to do it, we still have to haul off and do what it is we decided to do. Having made the decision by itself is not enough to force the onset of the action. Finally, there are actions which extend over fairly long periods of time. We decide to learn French or to swim the English Channel, and it is not enough to just begin the action, we have to make a continuous effort to reach the completion. Again, we experience a gap between the antecedent causes of the action, including the onset of the action, and the continuation of the action to completion.

I have talked as if there were three gaps, the gap of decision-making, the gap of the onset of the action, and the gap of the continuation of the action to completion, but in fact they are all parts of a continuous causal gap of conscious voluntary action. The experience of the gap is the experience of thoughts and actions where we do not sense the

decisions and actions as causally fixed by the antecedent conditions. It consciously seems to us that it is up to us what we decide and what we do. We do not have anything like this gap when it comes to digestion, or even perception. I can move my head to change my perception, but if I am looking at an object in broad daylight at point-blank range, it is not up to me what I see. I see what is there. The gap I have described is a feature of our conscious experience of deciding and acting in a way that it is not part of our conscious experience of perception, digestion, growth, or any number of other of life's events.

The problem of free will can now be stated succinctly: does the experienced gap correspond to anything in reality? The traditional name for the gap in philosophy is "the freedom of the will," but the fact that we experience this gap does not imply that the features it seems to represent actually exist in reality, that there really is a causal gap between the antecedents of our decisions and actions and the actual making of the decisions and the carrying out of the actions. So it is only because of certain sort of consciousness, the consciousness of the gap, that there is a problem of free will at all. Furthermore, let us suppose that free will is not an illusion, that there really is a causal gap or set of causal gaps of the kind I have been describing. Then free actions can only exist for a conscious agent, because it is only a conscious agent who has a sense of alternative possibilities in the gap. By the way, traditional psychoanalysis recognizes this fact in that it tries to get the patient to bring to consciousness his or her repressed motivations. The theory is that only when they become conscious can they be controlled by the agent. As long as the motivations are unconscious, they control us instead of us controlling them. There is thus a double connection between consciousness and free will. It is only because of a certain kind of consciousness that we are aware of the apparent phenomenon of free will at all. Even if we come to believe it is an illusion, it is only because we have a certain kind of consciousness that we have the illusion of free will. But furthermore, on the assumption that it is not an illusion, only a conscious agent can genuinely have free will, because only a conscious agent can make conscious decisions and perform conscious actions in the gap.

It is important to remember that when we talk about consciousness and conscious decision-making and conscious thought processes, we are talking about the subjective, logical, conceptual, and intentionalistic features of neurobiological processes in the brain.

Right now, as you read these words, there are electrochemical processes going on in your brain that have the semantic contents of which you are now aware. Now in our tradition, which is heavily influenced by dualism of mind and body, we are not used to thinking of brute, neurobiological processes such as neuron firings or the secretion of neurotransmitters into the synaptic clefts as having logical properties, but that is exactly what I want to insist on. The very semantic content that you are now aware of and the very semantic content that figures in those thought processes and decisions that we think of as exemplifying the problem of free will, are also neurobiological processes going on in the brain. One and the same sequence of events has both biological features and logical, conceptual and conscious features.

II: The Persistence of the Problem of Free Will

Free will is one of those issues in philosophy where it seems to me we have made very little progress in my life time. For many areas there has been remarkable progress—one thinks of the philosophy of language, political philosophy, moral philosophy, and the philosophy of mind as areas in which substantial progress is visible. But when it comes to free will it seems to me we are pretty much where we were fifty years ago. Why is that? The problem of free will arises because there is a conflict between two deeply held convictions and we do not see how to shake off either conviction. The first conviction is that human actions are natural events like any other events in the world. Human actions are part of the natural world as much as human digestion and human growth, along with the movement of tectonic plates and the growth of seeds into plants, and as such they are subject to natural forces. But this seems to imply that human actions are entirely determined, that they are as determined as any other biological process or for that matter any natural process in the world. No one supposes that the stomach or the liver have freedom of the will. Why should the brain be any different? Why should we suppose that brain processes give us freedom of the will any more than we would suppose that stomach processes give us freedom of digestion? Determinism seems to be overwhelmingly convincing. On the other hand we all have the experience of free decision-making and freely acting. We have an experience of making up our mind

between the alternative possibilities open to us, such that given a choice between action A and action B we choose action A but we know, or seem to know, that all other things being equal right then and there, we could have chosen action B. We cannot shake off the conviction of our own free will.

III. Rationality Presupposes Free Will

Someone might object to the claim that we cannot shake off the conviction of free will as follows: "All the same, free will might simply be an illusion like any other illusion. After all, we can't shake off the illusion of color even though many people agree with the scientific account that colors are an illusion created by differential light reflection striking our specific nervous systems. Why couldn't free will be a similar illusion?" If free will is an illusion, it is an illusion that we cannot shake off in the way that we might shake off other such illusions as rainbows, sunsets and even colors. If I become convinced that colors are a systematic illusion, I can organize my life in such a way that it can be consistent with the belief that colors are a systematic illusion. The problem of free will is special in that our ordinary, everyday actions require us to act under the presupposition of free will. If you are given a choice in a restaurant between the pork and the veal, you cannot say to the waiter, "Look, I am a determinist, I will just wait and see what I decide, I will just wait and see what happens." Even the decision to refuse to exercise your own free will, presupposes free will. Your refusal to exercise free will is intelligible to you as one of your own actions only on the assumption that you were acting freely. So even if you become intellectually convinced that determinism is true and free will is false, all the same you cannot act on this conviction. That is to say, you cannot treat your own voluntary action as something that just happens to you in a way that you can treat your digestive processes, or for that matter your visual experiences, as something that just happens to you. This is why, for example, we have a problem of the freedom of the will but not a problem of the freedom of perception or the freedom of digestion.

Though we cannot shake the presupposition of free will whenever we engage in voluntary decision-making, all the same that is no argument in favor of the reality of free will. It is just an important point about the distinction between free will and other features

of our experience such as the experience of colors, where we could readily grant that the experience of color, though unavoidable, is nonetheless a systematic illusion.

IV. Compatibilism is Not a Solution

So how are we to resolve this dispute? Many philosophers think it has already been resolved centuries ago by such people as Hobbes, Hume and John Stuart Mill. Their solution is called "compatibilism." William James called it "soft determinism." Compatiblists think that if we understand these notions correctly, free will and determinism are really compatible with each other. It is perfectly possible for all of our actions to be causally determined, and yet for some to be determined by such things as our own character or our own rational thought processes, and these are the actions we call "free". Are all of our actions determined? Yes, of course, as much as any other events in the natural world. Are some actions also free? Yes, of course, because they are determined by certain sorts of inner causes rather than external constraints. If, as Hume points out, I act under a threat, or if I am under some sort of coercion, then in such cases I do not act freely. But in the normal case where I am making up my mind whom to vote for or what to eat in a restaurant, I have a case of free action because the determination of my action has to do with my inner rational decision-making processes. So free will and determinism are compatible. The word *free* is opposed not to *caused*, or even *determined* but rather to *forced, compelled,* etc. To be free is to have such things as your own character and thought processes determine your actions. Of course those features of your character and your thought processes that determine your actions are themselves as much determined as anything else in the universe.

Does this really solve the problem of free will? I do not think it does. No doubt there is a use of the word "free" where "He acted of his own free will" is compatible with "The causal forces determining him to act were sufficient to fix the particular action that he performed rather than any other action. Only that action was causally possible in that situation." But that is not the sense of free will that really interests us in these discussions. The sense of free will that is important both in our own understanding of our selves and in our larger philosophical context is this: Are our actions such that the antecedents of the action are in every case causally sufficient to determine that that action had to occur and

not some other? Or are there some human events, specifically some human actions, which are such that the causal antecedents are not sufficient to determine that that action had to be performed? That even given all those causal antecedents, some other course of action was open to the agent? Compatibilism does not solve our problem, it just changes the subject to talk about the use of certain words. There is definitely a use of these words where it is compatible to say that the action was completely determined and yet it was free. For example, when people march in the streets carrying signs that say "FREEDOM NOW," they are not interested in the elimination of the laws of causation, they are typically interested in getting governmental authorities to impose fewer restrictions on them. But this use of words is not the one that troubles us when we are really worried about free will.

As I am construing the problem of the freedom of the will, if free will were true the world would be different from the way it would be if determinism were true. That is, I am supposing that it is not just a verbal question of how are we going to describe our actual behavior in a world of natural forces, whether we choose to describe some as free or not. Rather I am supposing that there is an empirical, difference between the world of free will and the world of determinism. What on earth could such a difference be? Let us explore this matter further.

The preferred way to find out whether or not the world is one way rather than another, is by a set of methods that, since the seventeenth century, we have come to call "scientific investigation." Could science solve the problem of the freedom of the will? I am now going to make some speculative efforts. I am going to try to imagine a scientific account that would support free will and one that would count against free will. Let me say at the beginning that from what we now know about how the world works, it seems overwhelmingly likely that determinism is true and free will is false. But we don't know a great deal about how the brain works, and what I am going to do is imagine how the brain would be different if free will were true from the way it would be if determinism were true.

V. The Readiness Potential is Not a Solution

First I want to digress to discuss the work of Deecke, Grözinger and Kornhuber[1] who seem to provide evidence in favor of determinism. Their experiments done over thirty years ago in Germany have in recent years been repeated and extended by Ben Libet[2] in San Francisco. Here is how the experiment goes. You tell a subject to perform some simple motor task such as, for example, pushing a button at random intervals. You tell him to make up his mind to push the button and to observe on a clock exactly when he decided that he was going to push the button. What was discovered in these cases is that two to three hundred milliseconds prior to his actually being aware of the intention in action to push the button, there is increased activity in his supplementary motor area. This increased activity is called "the readiness potential." There seems to be a neuronal antecedent to his apparently free voluntary action. Does this refute free will? Does the existence of increased activity in the supplementary motor area prior to the consciousness of a decision refute free will? Does it provide any evidence for determinism? Many people think that it provides decisive, or at least strong, evidence for determinism because an increased neuronal activity precedes the conscious decision. On their view, the brain makes up its mind to perform an action before we are consciously aware that we have made up our minds. So, to repeat the question, does the existence of the readiness potential refute free will? Even Libet admits that after the readiness potential has occurred, it is still possible for the agent to veto the action. That is, even given the presence of the readiness potential, the subject does not thereby have causally sufficient conditions for performing that action, because the subject, given this activity in the motor area, can still change his mind and decide not to perform the activity in question. It seems to me the best way to understand the implications of the readiness potential are contained in the very expression itself. Given the fact that the subject has a prior intention to do something at random intervals, the brain, so to speak, gets ready for him to do it, prior to his consciously initiating the action to do it. It is important to notice in these cases that he already has a prior intention to perform an action of the type in question. The prior intention manifests itself in his readiness to perform the action. The existence of the

[1] Deecke, Lüder, Grözinger, Berta and Kornhuber, H. H. "Voluntary Finger Movement in Man: Cerebral Potentials and Theory." *Biological Cybernetics*. Vol. 23. Springer-Verlag, 1976. pp. 99-119

[2] Benjamin Libet. "Unconscious cerebral initiative and the role of conscious will in voluntary action". *The Behavioral and Brain Sciences, 8*: 529-566 (1985).

readiness potential does not by itself show that determinism is true, nor for that matter that determinism is false. The question is still left open.

Furthermore, the examples used in the study of the readiness potential tend to be rather trivial examples of human behavior. What we are interested in is such questions as, "Was Churchill's decision in 1940 to continue fighting against the Germans a free action or was it determined?" This question can hardly be settled by finding out the level of neuronal activity in the supplementary motor area a few hundred milliseconds prior to his making up his mind. That is, the kind of actions that we are typically interested in are not such superficial actions as pushing a button.

VI. Can We Treat Free Will as a Scientific Problem?

Let us try to take the problem of free will and determinism as a scientific question, assuming that the existence of the readiness potential is not by itself a decisive solution to the problem. How should we proceed? So far, I have been writing as if it could be taken for granted that nature is normally completely determined. In this respect, I follow many other authors who write about the problem of the freedom of the will. But of course since the development of quantum mechanics over half a century ago we know that this is false. At the most fundamental level, at the level of the quarks and the muons, we know that there is an inherent randomness in nature, that nature is not completely determined and therefore not completely predictable at that level. We can only make statistical predictions at the quantum level. People often talk as if quantum indeterminacy were something that exists only at the micro level. But of course that is the most fundamental level, and the indeterminacies go all the way up. The point for determinism is that the indeterminacies tend to cancel each other out at the macro level, so we can treat the baseball and the baseball bat as Newtonian phenomena even though they are as much pervaded by quantum indeterminacy as any other part of nature. Determinacy is not the norm in the physical world. On the contrary, the physical world is pervaded throughout with indeterminacy.

It is important to emphasize how much the standard "scientific" conception of nature is inconsistent with what is known in physics. The standard conception, even among many professional philosophers is that the universe consists of very small entities

called particles and that these behave in a deterministic fashion described by "laws of nature". One consequence of this view is that if all of the particles were stacked exactly as they were at the time of the Big Bang, then the subsequent history of the universe would be absolutely identical with a history that has occurred so far. Given the same causes you would get the same effects. This is definitely not consistent with contemporary quantum mechanics.

Does quantum indeterminacy have any bearing on the problem of free will? Some people suppose it does because they believe that the absence of determinacy at the quantum level allows for free will at the higher level. But the problem with this is that the indeterminacy of the quantum level is a matter of randomness, and randomness is not the same as freedom. Random actions, the existence of random phenomena, does not offer any support for the existence of free, rational, decision-making, because free, rational, decision-making, though not determined, is nonetheless not random. Quantum mechanics seems to give us randomness and not freedom. Randomness may indeed be an absence of determinism, but it is not thereby a manifestation of the freedom of the will. Free will has to be something more than, or something quite different from, random events occurring.

I am reluctant to discuss the question of the freedom of the will in connection with quantum mechanics because it seems to me that when people talk about quantum mechanics without actually doing the mathematical physics, there is a high coefficient of sheer nonsense, and I am aware of that risk. The amount of hot air that is talked about quantum mechanics exceeds even that talked about the freedom of the will. However, in this article, we are allowing ourselves a certain amount of speculative liberty, so let us proceed.

Free will presupposes consciousness. Only for the conscious agent can there be such a thing as the freedom of the will. But as I emphasized earlier in this article, we know that conscious processes in the brain are sets of neurobiological processes that occur in the brain. Those processes have a higher level of description where they are described as subjective and qualitative and a lower level of description where they are described as electrochemical. Same processes, different levels of description. Now we know for a fact that the only parts of nature that are definitely nondeterministic are the quantum mechanical parts, so it looks as if, if there is any factual reality to the conscious

experience of non-determinism, that is to say freedom, there must be some connection between consciousness and quantum indeterminacy. So what is the connection between consciousness and quantum mechanics? There have been various attempts to explain consciousness in quantum mechanical terms, and some very distinguished thinkers, among them Stuart Hameroff, Roger Penrose and Henry Stapp have all sought quantum mechanical explanations of consciousness. It worries me, and perhaps it ought to worry them, that their work is not taken seriously by the most advanced researchers in neurobiology. This of course does not show that main stream neurobiology is right and their revolutionary view is wrong, but if I were a researcher I would be worried if the very best workers in the field had no sympathy with the line of research that I was pursuing. However, let us continue to pursue it and see where it leads us.

The fact that randomness is not the same as freedom used to seem to me an obvious and decisive objection against any introduction of quantum mechanics into the discussion of the problem of the freedom of the will. Quantum randomness does not give us anything even remotely approaching free will. Freedom requires that the agents' rational processes arrive at a conclusion in a way which is non deterministic, in a way which is such that the antecedent causes are not sufficient to determine the action in question, but where are the result is not in any sense random. So it looks like the randomness of quantum mechanics is simply irrelevant to the freedom of the will. However, it now seems to me that the argument as I just stated it commits a fallacy of composition. Here is how. The fallacy of composition is a fallacy of supposing that what is true of the elements of a system will be true of the entire system. So for example it would be a fallacy of composition to suppose that because neurons are firing at a rate of forty Hz that therefore the whole brain must be firing at forty Hz. That is a typical example of the fallacy of composition. Now the fact that the behavior of the sub neuronal particles is random, does not by itself show that the behavior of the whole system is random, even though the system is made up entirely of those particles. Certainly that is something we know from the behavior of larger systems generally. In theory it is possible that you might have a quantum indeterminacy at the level of baseballs and baseball bats, but though there is a mathematical possibility of random behavior of baseballs and baseball bats, in real life the indeterminacies cancel themselves out, and the trajectory of

the ball when the bat hits the ball can be calculated on the basis of deterministic Newtonian mechanics. So we might have randomness at the bottom level without having randomness at the higher levels. Now let us suppose that the hypothesis that consciousness requires a quantum mechanical explanation were true. I doubt very much that it is true, but at least it seems to me a possible empirical hypothesis. Here is how it goes: we do not have an explanation of consciousness. Standard accounts treat the neuron or neuronal groups, such as maps, as the fundamental explanatory level, but suppose that the minority is right in this case and that the explanation of consciousness will ultimately have to make an appeal to the quantum level. Suppose that it is only because of quantum phenomena, that the entire system made of neurons and synapses, which are themselves made of lower level atomic and subatomic particles, can be conscious. But now let us suppose that the following hypothesis is true. Suppose that the conscious decision-making processes inherit the absence of determinism of their sub atomic substrate, without inheriting the randomness. Let us suppose that the sub atomic particles with their quantum behavior are causally responsible for the existence of consciousness, but that certain forms of consciousness acquire an indeterminism which is not thereby random at the higher phenomenological level. In other words, it is a fallacy of composition to assume that the randomness at the bottom level must apply at the top level. It is at least logically possible that the absence of determinism at the bottom level could be inherited by an absence of determinism at the higher level without thereby inheriting randomness at the phenomenological level. This point has to be stated precisely. From the quantum mechanical view, the whole system will still be random, but at the quantum mechanical level we do not even have the vocabulary to talk about conscious decision-making. What I am suggesting is the logical possibility, though empirical unlikelihood, that the higher level consciousness of voluntary, free decision-making would manifest the lack of causally sufficient conditions characteristic of the quantum level without inheriting the randomness of that level. We would have the conscious experience of our reflecting and making up our minds having an effect on our behavior, without it being the case that the conscious reflection and decision-making and behavior were thereby totally random, nor totally fixed by antecedently causal sufficient conditions.

Notice that from the point of view of the atomic physicist, the randomness goes all the way up because he or she cannot make deterministic predictions based on causally sufficient conditions. All you can make are statistical probabilistic predictions. But here is where the idea that consciousness might have a quantum mechanical explanation comes in. We are supposing that the consciousness is real, and that it is important in determining the outcome of the conscious thought processes, but that it itself inherits the lack of causally sufficient conditions of the micro-level without thereby inheriting the randomness. The only sense in which it is random is that it is not predictable based on causally sufficient conditions. But then that is exactly the result that the free will theorist hopes to obtain.

I have to say that it is an extremely unlikely prospect that we could have conscious, rational decision-making manifesting the indeterminacy of the quantum level without becoming random at that higher level. In fact, the whole idea is intellectually uncomfortable. If it were right then we will have substituted three mysteries for one. I began by suggesting that free will is a mystery. We knew, before I began the talk, that consciousness and quantum mechanics are also mysteries. What I am now suggesting as a hypothesis is a system whereby the mystery of free will is explained by the mystery of consciousness, which is itself explained by the mystery of quantum mechanics. If that is not enough to give any respectable philosopher indigestion, then I think indigestion would not be a natural phenomenon.