Philosophical Explorations, Vol. 10, No. 1, March 2007

Routledge
Taylor & Francis Group

NEUROSCIENCE, INTENTIONALITY AND FREE WILL
Reply to Habermas

John R. Searle

I agree with much of Habermas's article 'The Language Game of Responsible Agency and the Problem of Free Will,' but concentrate on disagreements. (i) He is wrong to think the language game of neuroscience is somehow at odds with the language game of rational intentionality. I argue that they give different levels of description of the same system. He also has too narrow a conception of contemporary neurobiological research. (ii) He is mistaken in thinking there is a 'performative contradiction' in engaging in research that presupposes free will in order to disprove free will. (iii) His 'epistemic dualism' is irrelevant to the issue. (iv) He has some misconceptions about the world in general, especially about 'downward causation.' He seems to think that the physical world is deterministic. It is not. Quantum indeterminacy pervades the entire universe. We have the illusion of determinism because in some systems the quantum indeterminacies cancel out at the macro level. Is the brain a deterministic system? Right now we do not know.

KEYWORDS determinism; epistemic dualism; levels of description; neurobiology

I agree with much of Habermas' article, but in philosophy disagreements are usually more interesting than agreements. So most of this reply will be based upon articulating the points where I disagree.

1. Free Will as a Necessary Presupposition Even if Determinism is True

Habermas sees the problem of free will as essentially a problem involving a conflict between the scientific conception that we have of ourselves as determined parts of nature and our self-conception as free, rational agents.

'[*T*]*he problem of free will* presents itself as the question of whether the prospective progress in the neurosciences undermines' the language game of responsible agency (p. 14). I think that is only one aspect of the problem. I would put the problem in a way that I hope is consistent with Habermas but has a different emphasis: on the standard contemporary account of how the world works we are completely determined in all of our behavior and yet we cannot engage in rational decision-making processes or in voluntary intentional action except on the presupposition of free will. There is, in short, a flat inconsistency: on the one hand, we think all events are determined by antecedently sufficient causes, and

ISSN 1386-9795 print/1741-5918 online/07/010069-8

DOI: 10.1080/13869790601170169

on the other hand, we think that the antecedents of at least some of our actions are not causally sufficient to determine the action. It is up to us whether or not we perform the act. I put this in other writings (Searle 2006, chap. 1) by saying that we are aware of a *gap* between the causes of our actions in the form of reasons and the actual performance of the actions. Indeed in my view there are at least three phenomenologically real gaps: first between the reasons for an action and a decision to perform the action, second between the decision and the onset of the action, and third, where actions are extended over time, between the initiation of the action and its continuation to completion. The actual question concerning the freedom of the will is a question whether the phenomenologically real gaps in fact reflect the absence of sufficient causal conditions in nature itself. To put the question more precisely, is the brain a completely deterministic system or not?

If you believe you are determined you will find that you cannot live your life on the presupposition of determinism. For example, if asked by the waiter in a restaurant to choose which item from the menu you want to order you cannot say, 'look, I am a determinist. I will just wait and see what happens. Che sará sará.' Why not? Because that remark is only intelligible to you if you assume that its making was a free, intentional, voluntary performance on your part. The refusal to exercise freedom is intelligible to you only under the presupposition that it is a free action.

Having identified the problem, Habermas, in company with a very large number of other philosophers, sees its importance mainly in its relevance to the problem of moral and criminal responsibility. If you think that is the most important implication of the problem of free will you have not appreciated the seriousness of the problem. Moral and criminal responsibility are relatively peripheral aspects of the problem of free will. In a world in which we had decided to make no further use of the concepts of moral and criminal responsibility, the problem of freedom and determinism would still remain a desperate problem. To put the point succinctly, if determinism is true, it is not merely the case that we have no criminal and moral responsibility, but every single voluntary, intentional, conscious action of our entire lives was performed under a false presupposition. Why the fuss about criminal and moral responsibility if every time you raised your arm, drank a beer, got married, joined the Communist Party, chose chocolate over vanilla ice cream, enrolled in a university course, decided not to commit suicide, or did more or less anything at all, you did so under a false presupposition?

The problem is that we can only live on the presupposition of the freedom of the will, and yet, if the brain is a completely deterministic system, and our thoughts and behavior are entirely dependent on brain processes, then our lives are based on a false presupposition.

Habermas correctly sees that compatibilism is not a solution to the problem, and though I do not agree with his account of compatibilism I will say nothing more about it because we both agree that compatibilism does not solve the problem we are addressing.

2. Why There is No Necessary Conflict Between the Language Game of Neuroscience and the Language Game of Intentionality

My first major disagreement with Habermas comes when he construes the problem as essentially concerning two conflicting language games, the language game of neuroscience and the language game of intentionalistic explanations; and he then goes on the claim that the language game of science and the language game of intentionalistic

neuroscience are not only different language games, but they cannot even be made to connect with each other. He says, for example, that an argument involving conflicting reasons has to be judged by logical rules and cannot be described as a causal outcome of the states of the limbic system. Bodily states, he tells us, cannot contradict each other. He makes similar points at various places in the article. He says, for example, that the sorts of conditions that make actions intelligible are different in kind conceptually from the phenomena described by laws of nature. And indeed he seems to thinks that somehow or other it is *grammatically* impossible that the two perspectives he identifies should be described in a unified terminology that includes both mental operations and brain states.

I think he has an overly restricted conception of these language games and does not seem to realize that the two language games can simply be matters of different levels of description of one and the same system. The language games are not hermetically sealed in the way that he supposes. There is one level of description of my mental processes where they can be described as neurobiological processes in the brain. There is another level of description *of those very same processes* where they intrinsically have intentionalistic and semantic properties. Same processes, different levels of description. For example, my current desire that I would like another glass of wine is inconsistent with the conscious thought that I had better not have another glass of wine because I desire to be completely sober when I drive home. I consciously both desire to drink more wine and desire not to drink more wine. These are two inconsistent desires both realized in my brain in conscious neuronal processes. To say that brain processes cannot be inconsistent with each other is as mistaken as saying that sounds produced out of people's mouths cannot be inconsistent with each other. Speech acts, like conscious thoughts, have levels of description that do not identify their intentionality. But just as the speech act has a level of description where it is a sound, it also has a description where it has semantic properties; so the thought has a level of description where it consists of neurobiological processes, it also has another level of description where it has intentionalistic and indeed semantic properties.

Habermas has an overly narrow conception of contemporary neurobiological research. He thinks all neurobiologists proceed as if the brain were a simple mechanical system like a car engine. His picture is that, on the one hand, we have a self-conception as conscious, free, rational, thinking beings, and on the other hand, we have a conception of the brain as a complex physical system that operates on the same principles as any other machine. But in fact, though this did characterize neurobiology for a long time, a period in which most neurobiologists were reluctant even to approach the problem of consciousness, recent research is much more accommodating to the idea that the brain actually realizes consciousness, rationality, decision-making, etc. There are currently quite a number of investigations into consciousness and further investigations into selfhood, decision-making and rational action (Becchio, Adenzato, and Bara 2006; Jeannerod 2006; Searle 2005). Of course we have a lot further to go, and I would be the first to point out the limitations, but there is no obstacle in principle to having a neurobiological theory that treats the brain itself as the source and location of rationality and selfhood.

This, I believe, is a crucial flaw in Habermas's argument. For a long time most neurobiologists were reluctant to recognize the irreducibly mental, intentionalistic and conscious aspects of brain functioning. Like Habermas they thought that the 'language game' (not an expression, they used, of course) of neuroscience and the language game of conscious

rationality simply could not be made to connect, that consciousness, rationality etc. could not be construed as scientific problems. That is now very much changing. Consciousness has now become a major topic in neuroscience, and with consciousness certain other problems, such as the sense of selfhood and free action are becoming topics of scientific investigation. We know that mentalistic phenomena are real and irreducible, and we know that they are caused by and realized in neuronal processes. This means that if we are to understand them at the most fundamental level we have to understand their neurobiological base. Habermas would block us from that understanding *a priori* on the mistaken grounds that the two language games do not connect. He says 'natural scientific explanations exclude any inference to causally effective propositional attitudes (beliefs or desires)' (p. 22). I believe that is a deep mistake. There is no reason at all why we cannot have a neurobiological account of intentionality. Indeed the research pursuing such explanations is now going on (Becchio, Adenzato, and Bara 2006; Jeannerod 2006).

Habermas has an inaccurate conception of contemporary neurobiology. He repeatedly refers to its results as nomological and as finding law-like correlations etc., but if you look at any standard textbook of neurobiology you will be struck by the scarcity of 'laws.' What we are trying to do in basic neurobiological research is describe the processes by which the brain works. We of course suppose that all of those processes are grounded in the more fundamental phenomena of physics, but it is extremely unlikely that we will get law-like correlations between the phenomena we discover in neurobiology and the phenomena described at the micro physical level. It is extremely unlikely and so far virtually non-existent, that we will be able to find type–type correlations between the mental phenomena that interest us, such as memory and perception, and the phenomena of atomic physics. At this point we are not even sure what the right level of description of the brain is. Most textbooks assume that the neuron is the right functional unit, but it is by no means established that it is the right level at which we should be analyzing the phenomena. The main point of my criticism of Habermas here is that he does not seem to understand that it is an open question whether or not the brain is a completely deterministic system. Of course for methodological reasons we proceed as if it were, but that does not mean that is how it will turn out when we have a completely mature science of the brain.

3. Why There is No Performative Contradiction of the Sort Claimed by Habermas

My disagreement with Habermas about levels of description leads into my second important disagreement with him, about epistemic dualism and its relevance to the free will problem. What is epistemic dualism and why does he think it is even relevant to the problem? Habermas points out that we have two interacting stances toward the world. On the one hand, we have the concerned participant's stance where we are engaged in practical activities. On the other hand, we have the point of view of the detached observer. He points out that the scientist who claims to discover, from the detached theoretical point of view, that brain processes are entirely determined, must conduct his actual practice of scientific research from the concerned participant's stance, which proceeds on the presupposition of his own free will. Because the rationality of scientific investigation presupposes free will, Habermas thinks there must be what he calls a 'performative self-contradiction' in claiming on the basis of such scientific investigation that we do not have free will. I will make the following four comments on this argument, in ascending order of importance.

i. These are only two of a large number of possible stances that do not naturally fit into either of these categories. We can also take the aesthetic stance, the economic stance, the political stance, etc.

ii. It is a misuse of the notion of 'performative' to describe this as a case of a 'performative contradiction.' Austin, who invented the term 'performative,' would have shuddered at this use of the term. A performative contradiction would occur if I made a self-contradictory performative utterance. For example, 'I order you not to obey this order,' 'I promise you that I will not keep this promise.' Habermas's examples are not performative contradictions.

iii. I think it is a mistake to describe these two 'stances' as epistemic. The observer stance is one way of thinking of how things are independent of our activities, and the participant stance is one of thinking about what one is going to do and what one is doing. He points out correctly that in order to find out truth, which can be contemplated from the observer stance, we have to engage in the practical activity of investigation, which is conducted from the participant stance. He is apparently also aware that most of the time one is in both of these stances at once. So for example, when I ski down a mountain I worry about perfecting my technique (participant's stance), but I also reflect on the physics of the snow and the skis (observer's stance). There is nothing *epistemic* about my stances. I am not trying to find out anything, nor am I reflecting on the nature of my knowledge.

iv. There is no self-contradiction, performative or otherwise, in using the presupposition of free will to attempt to prove that we do not have free will. There is no problem in general in proceeding on the basis of a presupposition which, in the end, the investigation proves to be false. A scientist can investigate a domain *as if* something were true in order to prove that it is not true. An analogy will perhaps make this clear. A neuroscientist investigating color vision might presuppose the objectivity and reality of color and color discriminations as part of his research methods while at the same time concluding that colors are a systematic illusion. Similarly, a neuroscientist investigating freedom of the will might presuppose his own free will while conducting the investigation and nonetheless conclude that free will is a systematic illusion. There is no inconsistency in this whatever.

I hasten to add that I do not believe that free will is a systematic illusion. I do not know whether it is or not. But the objection that we have to presuppose our own free will when investigating the possibility of determinism does not settle the issue one way or another.

The bottom line of this discussion of the 'performative contradiction' and 'epistemic dualism' is that the performative contradiction is not a performative and not a contradiction and epistemic dualism is not properly construed either as dualism or as epistemic.

4. Why 'Epistemic Dualism' is Irrelevant to the Free Will Problem

But what I do not see is how this is even relevant to the problem of free will. Why does he think it is? Well he points out that the scientific investigator cannot treat his own investigations as if they were just natural phenomena, just events that occur in the world, he actually has to try to make rational decisions on the basis of evidence and criticism, and so on. Furthermore, these rational decisions can only be made on the presupposition of free will. But why is all of this even relevant to the problem? Yes, the naturalistic investigator cannot treat his own investigations as just determined events that occur, but must treat them as ongoing voluntary rational activities that he is both initiating and carrying out.

So what? This does not prevent the events that occur in the investigation from being determined events that occur.

Here is what actually goes on. The universe is huge and almost entirely meaningless. In at least one little corner, our tiny speck of a planet, there are conscious and intentionalistic beasts such as us, and these have created their own sorts of meanings. These beasts investigate nature and have different stances toward it. These stances do not in general affect the things they are stances toward. (Exceptions are observer relative phenomena, such as money or language, where the stance becomes part of the ontology. It is only money or a language if we think it is money or a language.) The existence of the two stances is so obviously irrelevant to the problem of free will that one wonders why Habermas thinks it is relevant. The only interpretation I can come up with is that he thinks the stances are somehow part of the ontology of free will. I suspect that the reason he thinks the dualism of stances is important is that he supposes these points of view are somehow or other part of the ontology of the phenomena that they are supposed to give us access to, as if somehow or other different points of view implied different realities. Why else, for example, would he make such a point of insisting that there is no 'view from nowhere'? Why is this even relevant? The fact that a view is always from somewhere implies nothing about the reality that is always viewed from somewhere. If he is supposing that his epistemic dualism has ontological implications, that the view becomes part of the reality viewed, then it is a very deep mistake, the deepest mistake in the entire article. Traditionally, this confusion of epistemology and ontology underlies idealism, and it is also very much a part of the phenomenological tradition.

Except for the little corner that we create by our stances, the real world does not give a damn about our stances; it just goes on as it is, including the stances, which are just another kind of natural phenomena. The mind is indeed part of natural history, and it is not a solution to the problem of free will to point out that often we take stances which treat it as if it were not part of natural history.

5. Habermas Has Some Misconceptions of Physics and its Relation to Mental Causation

I found some other apparent misconceptions and misunderstandings in the article. I think some of these may be serious. I will simply list them.

First, at one point he seems to endorse 'the perspective of a Laplacean demon, according to which there is, at any given moment, only one possible course of future states of the world. Accordingly, it is impossible for there to be two different worlds stemming from the same initial state' (p. 30). This view was widely held in physics for a long time. But with the advent of quantum mechanics it no longer characterizes physics at the most basic level. It is a consequence of quantum indeterminacy that the same causes could produce different effects on different occasions. Hence if all the molecules could be placed in exactly the circumstances they were at the time of the Big Bang, the subsequent history of the world could still be completely different.

Second, Habermas believes that the problem of free will has some special connection with 'downward causation,' and he even thinks that we need what he calls a better account of how downward causation can be made consistent with the principle of conservation of energy. Downward causation would be a case where the mental phenomena affect the physical phenomena of brain processes.

It seems to me quite obvious that mental phenomena can affect brain processes, and there is absolutely no question of violating the conservation laws in these cases. If, for example, someone says 'secrete acetylcholine at the axon endplates of your motor neurons,' I can do it by, for example, raising my arm. This is a clear case of downward causation. My intention in action caused the secretion of acetylcholine and it was my intention that it should. Of course the whole system only works because the so-called higher levels are grounded at the lower levels, at the level of neurons, synapses, and all the rest of it. In this respect the brain is like any other system that has many levels of description. The car engine can be described at the level of the spark plugs and the cylinder, but it can also be described at the level of the molecules of the metal alloys and the oxidization of hydrocarbons. The metaphor of upward and downward I think is misleading here. What we ought to realize is that the brain is a whole system, and it moves forward through time. The question of free will is whether or not that system is deterministic. If it is completely deterministic then we have no free will.

Third, related to the problem of downward causation is the problem of mental causation. Habermas erroneously supposes that on the account that sees consciousness and mental states as in some weak sense 'emergent' properties, there could be no mental causation. I think this is again a deep mistake. And again an analogy will perhaps make this clear. In some very weak sense of emergence, being a spark plug in a car is an emergent property of a collection of molecules, but this does not prevent the system functioning causally at the level of spark plugs, cylinders etc., just as it functions causally at the level of the passage of electrons and the oxidization of hydrocarbon molecules. These are not competing descriptions of two different systems; rather they are non-competing descriptions of one and the same system at different levels. Analogously the mind and consciousness function causally, though of course the causation of the mind is grounded in the brain. It is out of the question that the existence of free will should violate the conservation laws.

6. Conclusion

How then does all of this discussion bear on the problem of the freedom of the will? The problem of free will, to repeat, arises because we do not know how to reconcile two apparent facts about the brain with the situation in which we take ourselves to be conscious, rational, free agent. The two apparent facts are:

(i) All of our consciousness, including our conscious decision-making processes, is entirely dependent on lower-level brain processes. In the current jargon, they 'supervene' on brain processes, so there cannot be any feature of consciousness, intentionality, and all the rest of it, which is not completely accounted for by brain processes.

But at the same time:

(ii) On most standard contemporary accounts, including most neurobiological accounts, we assume that the brain is a completely deterministic system in the sense in which any other biological organ is a deterministic system.

But these two appear to be inconsistent with

(iii) We are free, rational agents.

The right way to approach this contradiction is to ask whether the two apparent facts really are facts. Specifically, we need to address the question whether or not the brain is a

completely deterministic system, and there it seems to me we don't know enough about the brain to have an answer to that question. I think it probably is, but the question remains an open empirical question not to be settled by philosophical analysis. If the brain is a completely deterministic system, then all of our behavior is determined and free will is a massive illusion. If the brain has an element of indeterminacy then, given certain assumptions about consciousness which I have tried to make clear elsewhere (Searle 2006, chap. 1), we have free will as a matter of empirical fact.

REFERENCES

BECCHIO, CRISTINA, MAURO ADENZATO, and BRUNO G. BARA. 2006. How the brain understands intention: Different neural circuits identify the componential features of motor and prior intentions. *Consciousness and Cognition* 15: 64–74.

JEANNEROD, MARC. 2006. *Motor cognition: What actions tell to the self*. Oxford: Oxford University Press.

SEARLE, JOHN R. 2005. The self as a problem in philosophy and neurobiology. In *The lost self: Pathologies of the brain and identity*, edited by Todd E. Feinburg and Julian Paul Keenan. Oxford: Oxford University Press.

———. 2006. *Freedom and neurobiology: Reflections on free will, language, and political power*. New York: Columbia University Press.

John R. Searle, Department of Philosophy, 314 Moses Hall 2390, University of California, Berkeley, CA 94720-2390, USA. E-mail: searle@cogsci.berkeley.edu