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

Routledge
Taylor & Francis Group

INTRODUCTION
Free will, neuroscience, and the participant perspective

Joel Anderson

Both within and (perhaps especially) outside academic philosophy, there is a powerful fascination with the idea that neuroscience is on the verge of demonstrating that free will is an illusion. Perhaps, it is said, we may be unable to live without this illusion, but there is no way to defend the notion that we are the authors of our actions without slipping into metaphysical dualism and anti-science. The authors in this special issue of *Philosophical Explorations* take up, from various angles, this suggestion that we must choose between science and free will, between naturalism and our sense of ourselves as agents. The central question is whether and how it is possible to fit free will, agency, and the mind into the natural world.

The aim of the lead article by Jürgen Habermas ('The Language Game of Responsible Agency and the Problem of Free Will'; published for the first time here) is to show that, at least as part of the 'language game of responsible agency,' free will is not merely compatible with science but that a complete, naturalistic account of the world must also make room for aspects of mind and agency that come into view only from the participant perspective. The 'space of reasons' must be integrated in our understanding of the natural evolution of the species. As Habermas sees it, this represents a departure from those versions of free will compatibilism that treat our commitment to the existence of free will as impervious to any discovery we could make about the world. Habermas is not satisfied with a sort of mutual non-aggression pact between neuroscience and common sense. What he is after is a genuine reconciliation of Kant and Darwin.

This is not exactly a modest goal, and so it is not surprising that there is some skepticism and dissension on the part of the commentators, Randolph Clarke, Michael Quante, John Searle, and Timothy Schroeder. The commentaries challenge Habermas's approach on a variety of counts, including the following: his analysis of the performative presuppositions of everyday actions, his attempt to distance himself from compatibilism, his downplaying of the threat that determinism poses to our self-understanding, his account of mental causation and the phenomenology of practical reflection, and his claim that an epistemic dualism based on fundamental differences between the participant and observer perspectives is both ontologically monistic and yet non-physicalistic.

This range of criticisms already provides an indication of the complexity of Habermas's argumentation, which draws on numerous interconnected aspects of his own systematic approach, including his earlier work on action theory, realism, pragmatics, and

ISSN 1386-9795 print/1741-5918 online/07/010003-9

DOI: 10.1080/13869790601170094

philosophy of science. In this introduction, I have set myself the task of filling in some of the conceptual background that is taken for granted in Germany (even among analytically inclined philosophers), a background that is still quite different from the background of philosophers raised on, say, Locke, Russell, Ayer, Kim, Lewis, Dretske, Fodor, and the Churchlands. Argumentative moves that seem natural (though perhaps still *mistaken*) to those with one background can be baffling and enthymatic to others. In what follows, I connect Habermas's position in the article published here to several themes that are developed elsewhere, especially those that tend to be less familiar. In the process, I shall highlight several of the points of disagreement with his commentators.

1. The Participant Perspective and the Performative Presupposition of Free Agency

Like most defenders of free will, Habermas starts out from our confident sense from ordinary experience that we exercise freedom of the will in weighing reasons, resisting urges, choosing how to act, identifying with or repudiating our motives (taking, that is a 'Yes-' or 'No-' position towards them), and so on. From inside this participant's perspective, the deterministic picture presented by neuroscience is difficult to recognize as an adequate picture of human agency. As a result, there is a deep tension between the participant and observer perspectives.

It is important to realize the ways in which Habermas is *not* appealing to our first-personal, everyday ('lifeworld') experience to argue for free will. It's not that determinism would be a truth we just couldn't bear (as Hume says in connection with skepticism), or that we would lose something of intrinsic value, in the sense of an impoverished ontology. The reason for rejecting determinism, for Habermas, is *not* that the costs of accepting the truth are too high. And the same goes for an approach to which Habermas is otherwise quite close, namely that spearheaded by Peter Strawson (1974; see also Wallace 1994), at least insofar as Strawson's approach is understood as suggesting that the presupposition of free will is so deeply rooted in our social practices and mutual expectations that relinquishing it would simply be too radical to pull off. And finally, it is not simply that the truth of determinism is inconceivable, in the sense of being very hard to imagine. Rather, Habermas is closer to Kant's original idea that one cannot conceive of oneself as a subject or agent if one views oneself as an object, and that conceiving of oneself as a subject is necessary for engaging in the sort of activity that supports our sense of having free agency.

This position is part of two broader aspects of Habermas's overall theoretic approach: his emphasis on the *performative* understanding of the first-person perspective and a commitment to the idea that not everything can legitimately be *treated as an object*.

In focusing on what agents presuppose 'performatively' in deliberating about and acting for reasons, Habermas gives the standard subjective-versus-objective contrast a decidedly Kantian inflection. The central idea is roughly that there are aspects of what it is to be an agent for which it is necessarily true that they can be reconstructed only on the basis of what agents must take to be true while they are engaged in acting. This is not a psychological claim, as Habermas understands it. In contrast to much of the free will literature the argument is not based on the *qualitative feel* of voluntary intentional action. Rather, it is based on what is, as it were, implicitly asserted in doing something, and on the internal tensions between such implicit claims.

It is in his work in the philosophy of language and, specifically, the pragmatics of communication that Habermas has worked out this account of what is performatively claimed.[1] In particular, it is central to his view that we raise various validity claims in (normal) speech acts. If, for example, I ask a student to open a window in a classroom, I am thereby implicitly 'raising validity claims'—in this case, that the window in fact opens (a claim to truth) and that it is appropriate for a teacher to ask a student to open it (a claim to normative rightness). I might be wrong on either count, of course, but even so my error would be about a claim that I have raised performatively. It is in this sense, then, that one can speak of 'performative self-contradictions,' which involve denying, either explicitly or implicitly (discursively or performatively), something that is presupposed performatively in making the utterance. Take the case of someone loudly uttering the sentence 'I am not now speaking.' The idea is that *either* we have not understood what the person is saying, *or* she is contradicting herself by denying explicitly what is presupposed, performatively, in making the utterance in the first place.[2] Moreover—and this is crucial for Habermas—there is no way to make that utterance without committing oneself to claims that are contradicting in the content of the speech act; it's unavoidable and necessary.

With regard to free will, any contradictions will be located deeper below the surface than many other paradoxes, partly because of the link to the notion of being a *participant* in 'the language game of responsible agency.' Habermas can be understood here as claiming that, just as being a participant in a tournament chess match requires a player to treat the piece of wood in front of her as a pawn and to attribute strategic motives to her opponent (otherwise, she is not really engaged in *tournament chess play*), there are also presuppositions associated performatively with one's genuinely participating in the language game of responsible agency. In this sense, according to Habermas, we can engage in the ubiquitous practice of being a responsible agent only if we assume that we (and others) are able to weigh reasons, judge on the basis of them, and allow our practical judgments to guide our actions. Note that, on Habermas's view, we are unable to dispense with these presuppositions not simply because of an internal contradiction within the individual, but primarily because of how our being able to *count* (even in our own eyes) as making a move within the relevant language game entails meeting *intersubjectively shared* expectations. This line of argument parallels Habermas's argument in the philosophy of language that being a competent speaker of a language is bound up with being able to recognize the assertability conditions for one's utterances, conditions that are essentially intersubjective in that they are not up to speakers themselves to decide (Habermas 1998).

The discussion of legal discourse on exculpating and excusing conditions then serves to illustrate one *institutionalized* microcosm of the more general language game of responsible agency, a context in which a great deal of work has been done to make explicit the mutual expectations about what counts as an appropriate 'move' in the language game or what sorts of moves require explanations. In addition, the discussion of legal discourse helps to clarify a point of frequent misunderstanding, as when readers of Habermas reject the presupposition on the grounds that people are not actually all that free, rational, and strong-willed (or, in the case of Habermas's philosophy of language, that they are not always oriented toward mutual understanding). Habermas's explicit position is that these are '*idealizing*' presuppositions: our practices make sense only on the basis of them, but we are entirely familiar—as are the legal courts—with the fact that we often fall short. Indeed, the questions of when the departure from the expectations provide grounds for excusing or exculpating someone are *themselves* part of the language game.

All four commentators have critical points to make about Habermas's claims regarding the performative presuppositions of the language game of responsible agency. One focus of criticism is his unwillingness to see his defense of the participant perspective as a form of compatibilism. Another set of questions turn on whether one couldn't allow all these practices to continue, while at the same time, in certain particularly philosophical moods or in the neuroscience lab, suspending these expectations that are tied to action and countenancing the possibility that the language game of responsible agency is actually just a game of pretend and that the real truth might lie elsewhere. And, for my own part, I believe that much more needs to be said (in connection with the notion of idealizing presuppositions) about the extent to which the demandingness of our expectations—the degree to which we *idealize* at the outset—ought to be adjusted on the basis of what neuroscience discovers about widespread limitations to our powers of will and cognition (for a parallel discussion, see Anderson 2001).

Before turning to Habermas's philosophy of science and his view that the world that is revealed exclusively from the participant's perspective is equally part of nature, I would like to mention briefly Habermas's profound, and profoundly *Kantian*, concern with the violations involved in treating persons as mere things. This is one of those themes where ontological and normative aspects are fundamentally entwined and it is a distinctive aspect of the Frankfurt School tradition of critical social theory with which Habermas is associated. This school is known for its opposition to the various ways of organizing social life (including some aspects of capitalism) that involve treating people as fungible, manipulable objects. This critique of 'objectification' and 'reification' owes a great deal, in the history of philosophy, to Kant and German Idealism, but it is also a normative notion that is central to many contemporary attitudes, as well as moral and legal principles. And this is why the debate over free will is so quickly felt to be a moral threat. It may not be *determinism per se* that has so many people worried—as Quante points out, the religious beliefs of millions of people involve accepting a version of predestination—but rather the implication this has: that we are the sorts of entities for whom determinism applies, namely, *things*. In this sense, Habermas's claim that there are 'limits to self-objectification' turns out to get some of its force from the concern that giving up the performative presupposition that our actions are up to us will end up undermining an invaluable normative distinction between what can be treated unproblematically as a mere object and what cannot.

There are, however, strong currents within neuroscience and philosophy claiming that we need to face up to the fact that we are, in the final analysis, mere things. Thoroughgoing physicalists will argue that there is no reason to think that giving up the person–thing distinction, *as an ontological watershed*, will leave us bereft of normative options. After all, we make distinctions about how it is appropriate to treat some objects and not others: slashing into a Rembrandt is wrong, but removing the wallpaper from an apartment's walls typically isn't. Since, from this perspective, the distinction between *being a thing as opposed to something else* isn't what's doing the work here, it is not clear why we can't do without it elsewhere. Persons may be special things, but we are still things. Kantians will, of course, have no truck with this, and so we come upon one of those very deep and momentous fault-lines where direct argumentation is likely to be unsuccessful and the most promising way forward is to see which side has the least unpalatable set of implications. And since it is frequently suggested that Kantian approaches to free agency and the mind have dualistic or unscientific side effects, it is not surprising that Habermas devotes so much of his article to diffusing this worry.

2. Philosophy of Social Science and 'Scientism'

Habermas's central strategy in defending his broadly Kantian approach is to go on the offense. As comes through perhaps most prominently in his reply to the commentators in the present issue, he sees 'scientism' as the real source of what is problematic in physicalistic approaches to free will (whether compatibilist or determinist). 'Scientism' is an uncommon term (although not, perhaps, an uncommon *view*), and the subject of some misunderstanding. Being opposed to scientism is not a matter of being opposed to science *per se* but of being opposed to a view whose exclusive focus on natural science models of explanation blinds it to both of the issues just discussed.

This is a theme that has occupied Habermas since his early work in philosophy of the social sciences (Habermas 1972, 1990; and the essays in Adorno et al. 1976), in which he defended interpretive or 'hermeneutical' approaches against models of social science that consider the only *real* explanations to be those that follow the basic principles of natural science. In some ways, much has changed within the philosophy of science since the heated 'Positivism Dispute' in 1970s Germany, especially in the wake of the 'post-empiricist turn' and an increased appreciation for interpretive approaches within the philosophy of social science and the idea that much of social reality can be grasped only in an intentionalist vocabulary or within the terms of a practice.[3] As these approaches emphasize, unless one is clued in on the practice, a soccer goal is just a ball hitting a net.

At the same time, as the current free will debate in Germany makes clear, among practicing neuroscientists and even much of the educated reading public, being 'scientific' is understood as a matter of seeking explanations that make no appeal to entities that do not admit of measurement and testing by the methods of the natural sciences.[4] And, indeed, as is clear from the exchanges in this issue, particularly between Searle and Habermas, the status of physicalism is very much at issue. In particular, if we assume that 'scientism' is understood as a pejorative term for a presumptuous and speculative ideology that fails to give the participant perspective its due and '(weak) naturalism' is understood as the form of ontological monism of anyone committed to the rational authority of natural science, then the debate turns out to be one over whether (or under what conditions) physicalism should be seen as more or less what naturalism calls for, or whether instead it is a form of scientism. Habermas is keen to defend 'weak naturalism' while rejecting physicalism as 'scientistic' in its blindness to the participant perspective, and he has devoted much of his most recent writing to developing his account of weak naturalism and the corresponding theories of reference and truth (Habermas 2003, 2005). But, as Quante's analysis shows, it is not always clear in Habermas's account what exactly counts as physicalism.

I won't try to resolve this issue or even summarize the positions taken. But it may be useful to highlight two related questions to which this gives rise.

One key question becomes, are there causally effective aspects of reality that are only accessible from within a mode of engagement that is fundamentally different from the third-personal, observer standpoint? It is in this sense that the question of mental causation—central to Schroeder's commentary here—is linked to the status of the participant perspective.

The other central question, as becomes clear in both Searle's remarks and those of Clark, is whether the reality of, say, soccer goals lies merely in the level at which it is described. Couldn't, in other words, physicalism allow for everything after all, simply by understanding the participant's perspective as one of the modes in which we gain access to something that is, in some ultimate sense, physical?

In a sense, both are questions about what, exactly, is supposed to make physicalism problematically scientistic.

3. Reasons-responsiveness and the 'Objective Mind'

For many free will theorists (and this is certainly true of Schroeder), the problem of mental causation is clearly at the heart of the free will debate. If actions are not ultimately explicable as merely the causal effects of brain events, it is suggested, it is hard to avoid the dualistic implications of saying that it was 'the agent himself' (as a distinct force) who generated the actions. Habermas's strategy to avoid this dilemma involves treating mental causation as essentially an interface between our brains (as a cognitive apparatus that has evolved to do just this) and our environment. Importantly, however, the environment is also cultural-symbolic, and this is what Habermas calls 'objective mind' ['*objectiver Geist*'] (see translator's note, p. 43), which includes the culture, language, institutions, practices, norms, and so on that structure and facilitate our thinking and acting. Mental causation, for Habermas, turns out to be a matter of our brains interfacing with this cultural domain. The important thing for his argument about the objective mind is that it is not expressible in a physicalist vocabulary and can be described only from the perspective of those who have been socialized into it, and yet it is also something which itself has developed and has its own *natural history*. In this respect, as well as in his defense of a 'weak naturalism,' Habermas' view shares some central tenets with McDowell's view (McDowell 1994).

The physicalist is likely to be suspicious and say, 'We know what drives are, and we know what it is to be aroused by or attracted to, say, the presence of food in the environment; but we have no idea what "reasons" are such that we could be responsive to them.' But think of the sounds that emanate from my mouth in saying, 'If you liked that Calvino novella, I think you'll really enjoy Don DeLillo's book *White Noise*.' For you to be responsive to this series of sounds as giving you a reason to do something, we need a very different model from that of how the presence of bananas in a tree gives me a reason to climb it.

Here it might be useful to compare this discussion with discussions about color perception. Within philosophy of mind, color perception offers an interesting challenge to physicalist theories, for it seems that, as a secondary quality, the color red is not itself present in red objects being perceived; rather, the redness is widely thought to be an observer-dependent fact. On views along these lines, at least, 'hyperintelligent beings' from a far-off planet might not have access to color, without it thereby being the case that color is 'merely subjective.' Whatever the best way to think about color-perception is, Habermas's point about the 'objective mind' can be understood as saying that there are aspects of human life—for example, that the piece of metal in my pocket is a one-euro coin, or that I'm two chess moves away from checkmate, or that I picked my daughter up early from school because she had a dentist's appointment—about which we can say not only that access to them depends on our being participants in the relevant practices but also that these practices are themselves the result of historical processes of development. Of course, since color perception is an evolutionary product, we can also say that redness has a natural history, although few would be inclined to say that the evolutionary changes in our perceptual system have altered the actual wavelengths involved. In the case of social reality, according to Habermas, *both* are in flux: whatever bio-mental-cultural apparatus parallels our perceptual system in giving us access to the social world, it is part of a

phylogenetic and ontogenetic process of development; and whatever the parallel with redness is—the art-historical significance of a particular painting or the 'space of reasons' itself—it too has a history as part of the natural world. In this way, the space of reasons is supposed to find a home in the natural world, thereby opening up ontological space for mental causation as a purely natural phenomenon.

4. 'Detranscendentalizing' both Science and German Idealism

This point about the history of the 'objective mind' introduces a new twist to the question of what it would be to have a genuinely full, complete account of the human mind and will—of our capacity to reason, explain, deliberate, and act. The classic German Idealist story of *Geist* [mind or Spirit] reflexively appropriating itself is a nonstarter for Habermas (though it has its adherents today); the natural world perceived from the objective perspective is not to be absorbed by *Geist*—the participant perspective absolutized, as it were. But Habermas considers the physicalist model of a *completed neuroscience* that would eventually absorb the mental on its own terms to be farfetched as well. In this context, it becomes clear once again that one of Habermas's central arguments against scientistic physicalism is that it is incapable of delivering the full account of the world it promises. For once physicalist natural science enthrones the observer standpoint as the final, freestanding arbiter of which claims about the world (and the humans in them) are true, it cuts off the possibility of explaining one important worldly phenomenon: *science itself*.

The 'detranscendentalizing' alternative Habermas is developing is ultimately to *naturalize neuroscience itself*, in the sense of 'weak naturalism.' The Darwinian insight, that a full account of the biological nature of an organism must also factor in the environment in which that organism's species evolved, finds its parallel in the idea that a full account of the mind (and the will) would similarly have to factor in the symbolic–cultural–social environment that has been an integral part of its emergence and development. This is what I take Habermas to mean when he speaks of a mind that could 'capture itself' [*sich einholen*]: the enterprise of providing a complete account of human agency and the mind would then not be exclusively a matter of identifying the requisite nomological regularities but also a matter of providing an account of how we came to be the sort of entities who can provide such an account. Habermas knows that he is making tentative speculations at the end of the essay regarding how such an enterprise might proceed and also knows how much work would need to be done to make this project go through, particularly in saying just exactly how to represent the unavoidable interlocking or meshing [*unhintergehbare Verschränkung*] of the participant and observer perspectives within such a complete account. But he is on familiar ground insofar as his is making a classic Frankfurt School point here, namely that our social practices and institutions—including practices of scientific research and explanation—are not timeless absolutes but historically contingent products of human activity.

Even if one grants the tentative nature of Habermas's discussion at the end of the essay, there are numerous objections that can be made. As Quante points out, there are questions that Habermas's formulations leave open, regarding the precise relationship between scientism and determinism, especially if—as Searle suggests—there is a real possibility of neuroscience discovering (as, in some sense, quantum mechanics already has) that the physical world is not entirely deterministic. If this is right, then there are problems with the opposition that Habermas is relying on, between that which is governed by natural laws and that which cannot understand itself as determined.

A further, and very pressing, issue has to do with the question of just what the difference is between Habermas's 'epistemic dualism' of interlocking perspectives and the multiple levels of explanation suggested by Clarke and Searle. The question is whether Habermas can insist that epistemic dualism is more than a matter of difference of levels of description, without thereby jeopardizing his commitment to ontological monism and weak naturalism.

As should be clear by now, Habermas's article approaches the free will debate in a number of genuinely innovative ways, bringing a wide range of theoretical considerations to bear on the issue. And if the initial responses from Clarke, Quante, Searle, and Schroeder are any indication, it is likely to generate a great deal of spirited discussion. I hope you enjoy this special issue on 'Free Will as Part of Nature.'

ACKNOWLEDGEMENTS

For helpful suggestions regarding this introduction, I am grateful to Pauline Kleingeld, Anthonie Meijers, and Marcus Willaschek. For assistance with the translations, I would like to thank Klaus Günther, Antti Kauppinen, and especially Jürgen Habermas.

NOTES

1. Most of the key texts are to be found in Habermas (1998, 2003, 2005).

2. As is clear from their exchange here, Searle prefers to reserve the term 'performative' to referring to 'performatives', in Austin's sense. Thus, 'On my definition, utterances of . . . "I am now shouting" (said in a loud voice) are not performative utterances' (Searle 2002, 159, fn. 3). Habermas has something different in mind, and it is only on the basis of his own usage that Habermas makes claims about performative contradictions.

3. For an excellent overview of the post-empiricist theories of Th. Kuhn, P. Feyerabend, M. Hesse, and others, see Bernstein (1983). See also Hiley, Bohman, and Schusterman (1992). See also the special issue of this journal, edited by Robrecht Venderbeeken and Stefaan Cuypers, on 'The Social Explanation of Action' (Vol. 7, No. 3, 2004). Finally, it is worth noting that this is an area where there is a great deal of agreement between Habermas and Searle; see, e.g., Searle 1997, 2002, chap. 8.

4. See, for example, Geyer's collection of newspaper articles (2004), and the two symposia in *Deutsche Zeitschrift fr Philosophie* (Vol. 52, Nos. 2 and 6, 2004).

REFERENCES

ADORNO, THEODOR W., R. DAHRENDORF, H. PILOT, H. ALBERT, J. HABERMAS, and K. POPPER. 1976. *The positivist dispute in German sociology*. Translated by G. Adey and D. Frisby. London: Heinemann.

ANDERSON, J. 2001. Competent need-interpretation and discourse ethics. In *Pluralism and the pragmatic turn: The transformation of critical theory*, edited by J. Bohman and W. Rehg. Cambridge, Mass.: MIT Press.

BERNSTEIN, RICHARD J. 1983. *Beyond objectivism and relativism. Science, hermeneutics, and praxis.* Philadelphia: University of Pennsylvania Press.

GEYER, CHRISTIAN, ed. 2004. *Hirnforschung und Willensfreiheit. Zur Deutung der neuesten Experimente*. Frankfurt am Main: Suhrkamp.

HABERMAS, J. 1972. *Knowledge and human interests*. Translated by J. Shapiro. Boston: Beacon.
———. 1990. *On the logic of the social sciences*. Translated by S. W. Nicholsen and J. A. Stark. Cambridge, Mass.: MIT Press.
———. 1998. *On the pragmatics of language*. Edited by Maeve Cooke. Cambridge, Mass.: MIT Press.
———. 2003. *Truth and justification*. Translated by Barbara Fultner. Cambridge, Mass.: MIT Press.
———. 2005. *Zwischen Naturalismus und Religion. Philosophische Aufsätze*. Frankfurt am Main: Suhrkamp. [English translation forthcoming in *Between naturalism and religion*, Polity Press.]
HILEY, DAVID R., JAMES F. BOHMAN, and RICHARD SCHUSTERMAN, eds. 1992. *The interpretive turn: Philosophy, science, culture*. Ithaca, N.Y.: Cornell University Press.
MCDOWELL, JOHN. 1994. *Mind and world*. Cambridge, Mass.: Harvard University Press.
SEARLE, JOHN R. 1997. *The construction of social reality*. New York: Free Press.
———. 2002. *Consciousness and language*. New York: Cambridge University Press.
STRAWSON, P. F. 1974. Freedom and resentment. In *Freedom and resentment and other essays*. London: Methuen.
WALLACE, R. JAY. 1994. *Responsibility and the moral sentiments*. Cambridge, Mass.: Harvard University Press.

Joel Anderson, Department of Philosophy, Faculty of the Humanities, Utrecht University, Heidelberglaan 8, 3508 TC Utrecht, The Netherlands. E-mail: Joel.Anderson@phil.uu.nl