Published in *Mind* 114 (2005)

Philosophical Foundations of Neuroscience, by M. R. Bennett and P. M. S. Hacker. Oxford: Blackwell, 2003. Pp. xvii + 461. H/b £65.00, P/b £19.99.

In this long and detailed book Bennett and Hacker set themselves two ambitious tasks. The first is to offer a philosophical critique of, what they argue are, philosophical confusions within contemporary cognitive neuroscience. The second is to present a 'conceptual reference work for cognitive neuroscientists who wish to check the contour lines of the psychological concept relevant to their investigation' (p.7). In the process they cover an astonishing amount of material. The first two chapters present a critical history of neuroscience from Aristotle to Sherrington, Eccles and Penfield. Chapter three (to which I shall return), offers the philosophical basis for much of the book. Chapters four to twelve present detailed philosophical criticisms of a wide variety of neuroscientists (and some philosophers) on a large number of topics. These include: Crick, Damasio, Edelman, Marr and Frisby on perception (particularly the primary/secondary quality distinction and the binding problem); Milner, Squire and Kandel on memory; Blakemore and others on mental imagery; LaDoux and Damasio on the emotions; Libet on voluntary movement; and Baars, Crick, Edelman, Damasio, Penrose, Searle, Chalmers, and Nagel on consciousness (with a great deal on qualia and self-consciousness). Chapters thirteen and fourteen, along with the two appendices, contain an elaboration and defence of the book's methodology and present explicit contrasts with the Churchlands, Dennett and Searle. Bennett and Hacker maintain that whilst neuroscientists have made significant discoveries concerning the workings of the brain, these discoveries have been obscured by their presentation within an incoherent conceptual framework. Their complaints, therefore, are often not with neuroscience itself but with what might be called its philosophical self image.

Published in *Mind* 114 (2005)

Two things must be made clear. The first is that this is, by the authors' own admission, a book for neuroscientists rather than philosophers. As a result, some of the philosophical claims that are brought to bear on neuroscientific theories are simply presented rather than robustly defended. This is understandable given the book's already considerable length. The second is that, as will be expected given Hacker's well known writings, the philosophical perspective from which neuroscientific theories are criticised is that of a Wittgensteinian philosophy of mind. Indeed, the book can be seen as an application of a certain Wittgensteinian conception of philosophical analysis to the domain currently investigated by prominent neuroscientists. It must be said, however, that even those unsympathetic to this Wittgensteinian approach will find that much of the critique sheds genuine light on the issues discussed. There are also long discussions of subjects that occupy philosophers in the analytic tradition. For example almost a third of the book is devoted to the topic of consciousness (one of the most interesting chapters being a sustained attack on the view that there is 'something it is like' to have an experience).

Bennett and Hacker cover an enormous number of topics and offer specific criticisms of a wide variety of neuropsychological theories. Of particular interest is the claim that the much discussed 'binding problem' is premised on a false view of perception as the seeing of an 'internal picture or image' (p.140), and the related criticism of Marr's computational theory of vision (both in chapter 4). However, by far the most frequent complaint is that neuroscientists succumb to what they dub 'the mereological fallacy'. Given its centrality, it is to this that I shall direct my attention.

The mereological fallacy involves attributing to a part of an animal that which can only meaningfully be attributed to the whole. In Bennett and Hacker's view, the mereological fallacy is rife in cognitive neuroscience, and is utterly pernicious. The

majority of neuroscientists discussed are found guilty, variously attributing to the brain (or to one of its hemispheres) the faculties of belief, knowledge, memory, perception and so on. As one of numerous examples, take Crick's assertion that, 'What you see is not what is really there; it is what your brain believes is there' (quoted on p. 68). Bennett and Hacker claim that the brain does not believe, know, remember, see, etc. anything. Rather, as they repeatedly claim, it is the person or the human being that is the only proper subject of such attributions. Claims such as Crick's are not false but incoherent, the result of a conceptual confusion that is, argue Bennett and Hacker, a direct descendent of the Cartesian misconception of ascribing psychological predicates to the mind rather than to the person. As a result, despite its professed materialism, contemporary neuroscience is presented within an incoherent framework that is fundamentally Cartesian.

The argument presented to support this charge of incoherence rests on a particular picture of the way in which we gain knowledge of our own and others' minds, and is associated with a version of the private language argument. It can be roughly stated as follows: we ascribe, say, pain to another on the basis of their behaviour (wincing and so on), but we do not infer the pain from that behaviour (p.81). Rather pain behaviour is 'a criterion' for being in pain (p.82), and gives us direct knowledge that someone is in pain (p.93). Such criterial grounds are partly constitutive of the meanings of psychological predicates (p.83). Given that brains do not, and cannot, display pain behaviour there is nothing that a brain could do that would satisfy the criteria for an ascription of pain (p.83). Thus, it is incoherent to say that a brain has, say, a pain.

This argument presupposes the falsity of an alternative picture of the mental. This holds that the meanings of psychological predicates are fixed by being internally

associated with one's own experiences. Bennett and Hacker argue that this would involve the postulation of a private language and that such a thing is not possible (p.97-100). This picture of an inner definition of the meanings of mental predicates is supported by the (false) view of self-knowledge as based on inner perception (p.91). A better way of understanding one's ability to self-ascribe psychological predicates is to say that psychological vocabulary, most obviously 'pain-utterances', is learnt as a substitute for natural expressive behaviour, and later as a tool for reporting one's states (p.101).

There is much here that I have glossed over, but the details will be familiar to many. The important point is that Bennett and Hacker's case rests in large part on this picture of psychological predicates as having their meaning fixed by types of behaviour that are 'criterial' evidence for their application. Unfortunately, this account of our knowledge of our own and others' minds is not sufficiently supported within the text. For example, the view that self-knowledge is based on inner-perception is dismissed in a single paragraph. The view that pain behaviour is 'conceptually connected' with pain is simply asserted: Bennett and Hacker claim that it is not, 'an intelligible possibility that pain might systematically be correlated with smiling and laughing' (p.82), but do not go on to justify this contention. Of course they may well be correct, but given that so much of the argumentative structure of the book relies on this point it cannot simply be stated as an 'evident logical feature' (p.83). Of course, as mentioned above, this is a book for neuroscientists. As such, the presentation of subtle philosophical arguments is of necessity brief (a more detailed account can be found in P. M. S. Hacker, *Wittgenstein: Meaning and Mind*, Blackwell: Oxford, 1990), but the philosophically curious reader will rightly demand more.

The book carries out its dual tasks within a strict division between the a priori and the empirical; between philosophical analysis and scientific investigation. Indeed, one of Bennett and Hacker's most frequent complaints concerns the disregard shown to this distinction by neuroscientists and philosophers alike. As they see it, the relationship between conceptual and empirical issues is one-way. Whilst it is important that philosophers clarify the concepts used in neuroscience, it is a mistake to think that neuroscience could have much of an impact on philosophy (other than to create new philosophical puzzles). They go so far as to say that, 'the supposition that scientific evidence may contravene a philosophical analysis is…risible' (p.404). This seems somewhat problematic given that, as part of their account of the meanings of psychological predicates, Bennett and Hacker offer an empirical story as to how psychological vocabulary is learnt. They claim that self-ascriptions are learnt, not by associating names with sensations, but as replacements of natural expressive behaviour, and they offer an account of how this might be achieved via parent-child interaction (pp.100-103). But this is precisely the sort of account that one would expect to be supported (or otherwise) by an empirical account of language acquisition. Far from being a mere platitude, the account is presented as being incompatible with (or at least evidence against) the introspectionist account of self-knowledge. It seems that the division between the philosophical and the empirical allows for some interplay after all.

In a short review it is only possible to scratch the surface of such a long and ambitious book. Bennett and Hacker do expose a great deal of unclarity in the writings of several leading neuroscientists, and do so in a way that displays a serious knowledge of the field. It also represents a challenge to many orthodox philosophical views and is always both provocative and entertaining.

Published in *Mind* 114 (2005)

JOEL SMITH

Department of Philosophy
University of Essex
Wivenhoe Park
Colchester
Essex
CO4 3SQ