Philosophical Foundations of Neuroscience

SECOND EDITION

M.R. Bennett
and
P.M.S. Hacker

WILEY Blackwell

Reviews of *Philosophical Foundations of Neuroscience*:

'This remarkable book, the product of a collaboration between a philosopher and neuroscientist, shows that the claims made on behalf of cognitive science are ill-founded. The real significance of impressive recent developments in the study of the brain, they allege, has been clouded by philosophical confusion in the way in which these results have been presented. The authors document their complaint in a clear and patient manner. . . . They disentangle the confusions by setting out clearly the contrasting but complementary roles of philosophy and neuroscience in this area. The book will certainly arouse opposition. . . . But if it causes controversy, it is controversy that is long overdue. It is to be hoped that it will be widely read among those in many different disciplines who are interested in the brain and the mind.' *Sir Anthony Kenny, President of the British Academy (1989–1993)*

'Overall the book provides the most thorough critical survey of the ruling theories of mental phenomena as they figure in contemporary science. The attention to detail is meticulous, and the philosophical analysis outstandingly lucid. Contemporary scientists and philosophers may not like Bennett and Hacker's conclusions, but they will hardly be able to ignore them. The work is a formidable achievement.' *John Cottingham, Professor of Philosophy, University of Reading*

'Contemporary neuroscience is an exciting, ebullient field and its practitioners are not much given to self-doubt. This dissection of the field by Bennett and Hacker ought to provoke some misgivings. Arguing for a sharp distinction between conceptual analysis of our everyday psychological concepts on the one hand and empirical, neuroscientific investigation on the other, Bennett and Hacker conclude that many neuroscientists – and some of their philosopher friends – have ignored or muddied that distinction at their peril. In particular, they argue that the misuse of psychological concepts in the interpretation of neural processes does not lead to testable or even false claims, but to nonsense. Neuroscientists, psychologists and philosophers will be challenged – and educated – by this sustained and well-informed critique.' *Paul L. Harris, Professor, Human Development and Psychology, Harvard University*

'[It] will certainly, for a long time to come, be the most important contribution to the mind-body problem there is.' *G. H. von Wright (1916–2003), Research Professor, Academy of Finland and Professor of Philosophy at Cambridge, Cornell and Helsinki*

'Sweeping, argumentative, and brilliant, this book will provoke widespread discussion among philosophers and neuroscientists alike.' *Dennis Patterson, Notre Dame Philosophical Review*

'Devastating critiques of psychologists and neuroscientists. . . . Whether this book leads to a reconfiguring of contemporary neuroscience and the philosophy associated with it will tell us much about the dynamics of contemporary intellectual life.' *Anthony O'Hear, Philosophy*

'This book is a joy to read. . . . A model of clarity and directedness . . . [Bennett and Hacker] have produced that rarity of scholarship, a genuinely interdisciplinary work that succeeds. . . . This is a wonderful book that will illuminate, provoke and delight professional scientists, philosophers, and general readers alike.' *Damian Grace, Australian Book Review*

'Clinical precision and . . . relentless good sense . . . [a] thoughtful and useful treatise.' *Daniel N. Robinson, Philosophy*

'Mandatory reading for anybody interested in neuroscience and consciousness research. The vast spectrum of material in philosophy and neuroscience that Bennett and Hacker consider is impressive and their discussion is thorough and illuminating.' *Axel Kohler, Human Nature Review*

'A delicious cake of a book in which Bennett and Hacker guide the reader through a conceptual minefield of confusions repeatedly made by neuroscientists and philosophers alike.' *Constantine Sandis, Metapsychology*

'Anyone who has ever framed a theory or explained one should read this book – at the risk of forever falling silent.' *The Rector, University of Sydney, Obiter Dicta*

'Impressively lucid . . . Bennett and Hacker unquestionably succeed in challenging our own concepts, examine them for dross, and strive to home in on fundamentals.' *Neil Spurway, Journal of the European Society for Study of Science and Theology*

'The fruit of a unique co-operation between a neuroscientist and a philosopher. . . . An excellent book that should be read by all philosophers of cognition and all researchers in the cognitive neurosciences.' *Herman Philipse, ABG #2, De Academische Bockengids*

'This book is an intellectual delight to read, whatever one's opinions on the subjects discussed, and it is impossible not to learn from it. Whether cognitive neuroscience is an appropriate method for a scientific psychopathology is an important question for psychiatry in the twenty-first century and this book raises critical issues in indirectly addressing this question and as such is important reading for psychiatrists, cognitive neuroscientists, psychologists and philosophers.' *Matthew Broome, International Review of Psychiatry*

'[T]here are, I think, grounds for hope that this book will do an enormous amount of good, both in correcting philosophical confusion within neuroscience and in promoting a new style of dialogue between neuroscience and philosophy.' *David Cockburn, Philosophical Investigations*

'Filled with pedagogical and constructive advice, this substantially new edition is the *catalogue raisonné* that many practising neuroscientists and neurologists had been calling for. By expounding them solely on their own merits, it makes clear why many of the propitious discoveries and theories about brain and mind that we live by shine imperial and how they – just as often – reveal themselves disquietingly unclothed.' *Juan M. Pascual, Professor in Neurology, Physiology and Pediatrics, University of Texas Southwestern Medical Center*

'No single work of neuroscience has a greater bearing on all others, or higher warrant to be read ahead of them. A peerlessly incisive analysis, ranging far across the domain, it lays bare the conceptual bedrock on which scientific enquiry rests and the claims of neuroscience ultimately stand or fall. It exposes a litany of errors that render empirical questions unanswerable by robbing them of the sense both truth and falsehood presuppose. It shows that though conceptual these errors have material, real-world consequences irremediable by empirical effort alone.' *Parashkev Nachev, Professor of Neurology, Institute of Neurology, UCL*

'The first edition of *Philosophical Foundations of Neuroscience* was essential reading for cognitive clinical neurologists, and two decades later a second edition, which surveys and criticizes the conceptual forms of developments since the turn of the century, is most welcome.' *Martin Rosser, Professor of Neurology, Institute of Neurology, UCL*

Philosophical Foundations of Neuroscience

Philosophical Foundations of Neuroscience

M. R. Bennett and P. M. S. Hacker

Second Edition

WILEY Blackwell

This edition first published 2022

Edition History
Second Edition © 2022 John Wiley & Sons Ltd.
First Edition © 2003 M.R. Bennett and P.M.S. Hacker, Published by Blackwell Publishing Ltd.

Registered Offices
John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

Editorial Office
111 River Street, Hoboken, NJ 07030, USA

For details of our global editorial offices, customer services, and more information about Wiley products visit us at www.wiley.com.

Wiley also publishes its books in a variety of electronic formats and by print-on-demand. Some content that appears in standard print versions of this book may not be available in other formats.

Library of Congress Cataloging-in-Publication Data
Names: Bennett, M. R., author. | Hacker, P. M. S. (Peter Michael Stephan), author.
Title: Philosophical foundations of neuroscience / M.R. Bennett and P.M.S. Hacker.
Description: Second Edition. | Hoboken, New Jersey: John Wiley & Sons, 2022. |
First Edition: 2003. | Includes bibliographical references and index.
Identifiers: LCCN 2021008649 | ISBN 9781119530978 (Paperback) |
ISBN 9781119531012 (PDF) | ISBN 9781119530633 (ePub) | ISBN 9781119530923 (eBook)
Subjects: LCSH: Cognitive neuroscience--Philosophy.
Classification: LCC QP360.5 .B465 2022 | DDC 612.8/233--dc23
LC record available at https://lccn.loc.gov/2021008649

Cover design by Wiley

Set in 10/12pt BemboSTD by Integra Software Services Pvt. Ltd, Pondicherry, India

10 9 8 7 6 5 4 3 2 1

For Gillian and Jocelyn

Contents

Foreword to the Second Edition

Denis Noble CBE FRS hon FRCP

The publication of the second edition of this book nearly two decades after its first publication is a suitable occasion to review what it achieves and why that is important.

It has certainly succeeded in bringing a high degree of rigour to the interaction between science and philosophy in the field of neuroscience. Many of the questions raised by scientific discovery are conceptual and cannot be answered by further empirical discovery alone. Nor can conceptual analysis be dissociated completely from empirical discovery. As just one example, the deep questions about the nature of our universe raised by the discoveries of quantum mechanics and relativity would not have seemed relevant if nineteenth-century certainties about a purely deterministic universe working in a purely Cartesian space had been confirmed. That is one of the reasons why collaboration between active scientists and active philosophers is necessary.

It is also one of the reasons the authors refer in their introduction to 'the fact that the potentiality for conceptual confusion is buried deep in our language. Such confusions can be eliminated for a few decades by painstaking conceptual analysis. But they will rise again, as younger generations fall into the same traps. Sense data died under critical onslaught in the 1950s and 1960s, but by the end of the century internal representations arose phoenixlike from their ashes.'

It seems to me to be obvious that language needs constant re-analysis as the meanings of words change, new metaphors arise and new potential confusions occur. Yet, by and large, twentieth-century science was not ready to accept that philosophy had anything of any importance to contribute. That view was based on the idea that science and philosophy as they were understood in the seventeenth century had confused the two, even to the extent of naming the first scientific journal *Philosophical Transactions of the Royal Society*. Originally published in 1665, its first editor, Henry Oldenburg, was as much at home discussing (in long correspondences in Latin) with the philosopher Benedict Spinoza as he was with the scientist Isaac Newton. One of Spinoza's great philosophical works was nearly published in the journal.

In later centuries the idea grew that, once issues that had initially been raised as philosophical issues had become the subject of practical empirical enquiry, there was no longer any need for further philosophical analysis. That depended of course on the conviction that the initial conceptual distinctions had been set in stone and were no longer open to question.

I see the signs that the twenty-first century is proving to be more aware of the pitfalls this creates. To take just one example that has been the subject of my own research recently, the discoveries that led to the so-called central dogma of molecular biology, formulated by Crick in 1958 after the earlier empirical discovery of the double helix, were presented in the last century as an unquestioned

empirical fact. Yet the reason they were viewed in that way was itself based on a deep misunderstanding of the nature of the DNA molecules. Following in the footsteps of the quantum-mechanics pioneer Erwin Schrödinger, the genetic material was assumed to reproduce itself like a crystal. You will find that assumption hidden away in the textbooks, and sometimes openly acknowledged in the popularizations, such as Richard Dawkins's *The Selfish Gene*, where he explicitly says, of DNA replication, that 'This is how crystals are formed.'

We now know that DNA does not function like a crystal in living cells, nor does it reproduce itself accurately. In fact, the copying process is so inaccurate that there would be hundreds of thousands of copy errors if the cell did not come in to ensure faithful transmission to the next generation by systematically proof-correcting the inaccurate copies.

With that fact alone, many other foundations in evolutionary biology turn out to be conceptual errors. I have detailed those errors elsewhere.[1] They are fundamental to our view of ourselves as humans and our place in the universe, and raise many philosophical questions that had been considered closed, such as whether we and other organisms are purposeful.

Neuroscience, like any other field of science, cannot be immune from such problems raised by assumptions that creep into our views of the world and then become treated as accepted facts. The problems raised by metaphysical assumptions masquerading as empirical facts are just one example where collaboration between science and philosophy is necessary.

Finally, I wish to draw attention to the fact that this edition is not simply the original book updated. As the authors explain in their introduction, the book has been substantially rearranged to separate out conceptual problems that individually require more extensive treatment. Moreover, a vast literature, particularly on new technical methods, is out there to be taken into account.

I particularly appreciate the fact that there is now a separate chapter (3) concerned with the conceptual problems arising from ascribing to the brain properties that can only sensibly be ascribed to the organism as a whole. I see this point as a natural ally to my own arguments for multilevel interpretations of organisms (the principle of biological relativity), since those arguments lead to demonstrations that it does not make sense to ascribe functions, purposes and goals to levels of organization that could not possibly integrate those functions. As an example from my own field of physiology, it does not make sense to ascribe to the molecules of the heart the function of pumping blood around the body. Functions and purposes can only be ascribed to levels where they make sense (in this case to the complete circulatory system), and some of those (psychological attributes) are necessarily applied sensibly only to the whole living being. As the authors emphasize throughout their book, if a property cannot sensibly be ascribed to something, then it is not an empirical question whether it is or is not the case. Both answers would be meaningless.

Separating chapter 3 from its related conceptual problems, such as introspection, enables those problems to be more thoroughly analysed in chapter 4. While reading that chapter I tried to imagine what it would be like for me to be in the privileged position of being an observer in some future neuroscience laboratory. My privilege would be to observe, through some yet-to-be-invented high-resolution process (vastly higher resolution than current scanning methods), the detailed molecular and electrical neural and other body processes that had been discovered to be those associated with me thinking about, for example, the square root of minus 1. And I realized of course that, were that to be possible, I would be just like those neuroscientists. I would have no idea how to interpret all the electrical and molecular events as somehow 'being' my idea of the square root of

---

[1] D. Noble, 'The illusions of the modern synthesis', *Biosemiotics* (2021), https://doi.org/10.1007/s12304-021-09405-3.

minus 1. I would be just as ignorant as I would be if gazing at the series of 0s and 1s in a computer readout of its binary-number activity when calculating a problem involving complex numbers.

That inability in understanding my own brain processes would have nothing to do with the problem I already had as a student when first grappling with and learning the concept of imaginary numbers (I use this example only because it readily shows just how absurd it would be to claim that one could 'see' imaginary numbers in my brain!). For I would also be none the wiser if the question I had been imagining while my brain was being examined was a much less problematic one, perhaps what I wished to eat for breakfast. The only way for those processes to be understood, by me or by the neuroscientists, would be for it to *be* me thinking those thoughts and telling the world what I was thinking. But the neuroscientists could learn that directly from me without recording from my brain.

I think this is a general problem in the multilevel understanding of organisms, not limited to the brain and nervous system. As I have already indicated in this foreword, all science requires conceptual analysis as an ongoing process.

Denis Noble CBE FRS hon FRCP
31 May 2021

Foreword to the First Edition

Denis Noble CBE FRS hon FRCP

This book was simply waiting to be written. The reductionist agenda in biological science has generated so many conceptual difficulties that someone, sometime, had to analyse these problems in depth from outside the reductionist viewpoint. That a neurophysiologist and a philosopher should combine to do so is also a sign of the times. As biology moves on to address the complexity and extraordinary subtlety of life, now that it has broken it down into its smallest pieces, we will find this kind of combination of skills and ways of thinking even more necessary. As the authors make clear, philosophy (at least in the analytical form practised here) and empirical science are not in opposition. Rather they deal with different kinds of question. Yet, since a conceptual scheme is necessary to any fruitful experimentation, we cannot avoid asking both kinds. Keeping a clear head while we do so is not as easy as it may seem!

I must issue a warning: this book is *highly* controversial. Some of my scientific colleagues will strongly challenge, and will surely be deeply provoked by, the claim that neuroscience has frequently and systematically confused conceptual and empirical questions. To them I would say, first, that the authors clearly recognize the brilliance and phenomenal achievements of the scientists whose conceptual work they analyse. This is emphatically not a book debunking experimental science, any more than the fact that most physiologists now dismiss the dualist philosophy of Sherrington or Eccles detracts in any way from recognizing the immense significance of their scientific achievements. We find it perfectly possible to admire the experimental and associated analytical skills while wincing when we see how completely trapped they were in their outdated and indefensible philosophical position.

Second, I would appeal for some patience and humility. Patience, because as a physiologist who has interacted with (and published with) professional philosophers of various persuasions for over 40 years,[1] I have to say that I find scientists unthinkingly debunking philosophy more often than the other way round. Humility, because the issues are of the utmost social importance. Some of the claims of reductionist science are not only conceptually incorrect or even unintelligible, they have major social implications. The words we use, the concepts by which we analyse and present biological discovery, deeply affect the way in which we see ourselves as human beings. For that reason, if for no other, a critical debate is necessary. The authors of this book have thrown down a major challenge in that debate.

---

[1] See, for example, D. Noble, 'What do intentions do?', in A. Montefiore and D. Noble (eds), *Goals, No Goals and Own Goals* (Unwin Hyman, London, 1989) and D. Noble, 'Biological explanation and intentional behaviour', in W. H. Newton-Smith and K. Wilkes (eds), *Modelling the Mind* (Oxford University Press, Oxford, 1991).

The controversial nature of this book arises because the particular reductionist philosophical position it criticizes is very widely held today within the scientific community (and also by some well-known philosophers). Moreover, for most of them, this position is a *methodological* necessity, perceived to be the only paradigm for science to successfully explain things. The first reaction to the counter-argument, as presented here, will be to protest that somehow science is being (unnecessarily?) circumscribed; that some problems are, as it were, being taken from its grasp. I would argue the other way. The first step to scientific progress is to ask the right questions. If we are conceptually confused, we will ask the wrong questions. The authors illustrate this in detail with many examples.

It is hard to escape the confines and confusions of the culture in which one finds oneself. The history of philosophy shows that, just as much as the history of science. The central appeal of this book is to throw off the remaining legacy of the Cartesian confusions, first expressed as a duality of mind and body, but latterly expressed as a duality of brain and body. The authors show that, although the first required belief in a non-material substance, while the latter is wholly materialistic, many of the conceptual problems (essentially those of the 'ghost in the machine') are the same. For our dualist predecessors the ghost was an actual immaterial substance, for us it is 'the "I"' (or 'inner eye' or whatever) that 'sees' the qualia that 'form our experience'. This is what may lead us to ask which group of cells, or even which neurone(!), is doing the 'seeing'. The point here is that simply replacing 'I' or 'inner eye' by the brain or a part of the brain doesn't avoid the problem.

The key to understanding the confusions here lies in an analysis of the logical conditions for ascribing mental and psychological properties. This is not easy. It involves one of the most difficult of twentieth-century philosophical ideas, that of the 'private language argument': what it is to say things like 'I feel pain' or 'I see red'. I struggled through the ramifications of this argument many years ago before writing my own contributions to the philosophy of biology. I wish I had had the benefit of the relatively easy path that Bennett and Hacker have provided. Even those who fundamentally disagree with their arguments (and I look forward to seeing them engage in debate) must surely acknowledge that this is a sustained and valuable exposition of an important and influential philosophical position.

Although I would describe that position as philosophically radical (in the correct sense of that word: going back to basic roots and eradicating those that shouldn't be there), it is often dismissed by scientists as conservative because it may appear to restrict using language in new ways. Yet, they would argue, science cannot advance without doing that. And what better way to achieve it than to start with metaphor or *façons de parler*, consolidate with dead metaphor (metaphors that become part of everyday language – constructivists argue that that is the way language evolved) and finally end up with a change in our conceptual scheme? Indeed, why not, if that is what will enlighten us, lead us into new conceptual territory, formulate new theories. But there is a simple test for whether that could work in any particular case. For each such metaphorical (or similar) change in use or meaning, or novel piece of terminology (such as 'qualia' or 'memes'), imagine stating its opposite, and then ask whether any conceivable experiment could test empirically between the two. The deep problem for many 'novel' concepts and language uses in reductionist approaches is that this test totally fails. The novel use of language is then not so much a scientific as a political or social tool. If you doubt this, try imagining an experiment to test between the existence or non-existence of qualia. Or for whether or not the brain makes representative maps (which are *not* homunculi incidentally). Or for brain states that 'explain' rational thought (rather than being a necessary physical basis for its existence). Surely we should only introduce new terminology where, as with quarks and black holes, we provide the empirical criteria for determining their existence?

Perhaps the problem for many scientists is to imagine what would happen if we abandoned the universality of the reductionist approach. For sure, the nature of science would change. But so it should! We would have to recognize that causation and explanation do not always run upwards from lower to higher levels. And, surely, at a time when we have already come to understand the

extent to which causation runs in the opposite direction (higher-level states in biological systems even influence something as fundamentally lower-level as gene expression), how can we possibly imagine that we will progress without recognizing the validity of explanations at all levels? One of the criteria for determining the level at which explanation succeeds is to ask what can sensibly be ascribed at different levels. It does not make sense to look for explanations at levels lower than that for the applicability of the relevant predicates.

This is particularly true of rational behaviour, including the use of language. The argument is basically very simple. We cannot, coherently, deny our own rationality. Otherwise we would have difficulty meaning what we say or being convincing in saying it, which is precisely what happens in the sad cases of those mentally ill people who nevertheless are aware of, but can't help, their irrationality. If we really could succeed in 'reducing' rational behaviour simply to molecular or cellular causation then we would no longer be able meaningfully to express the truth of what we had succeeded in doing. But, thankfully, no such reduction is conceivable. We know what it is to be rational, and what it is to lose that capacity. That knowledge has nothing to do with the question whether there exist specific and causally sufficient neural states and interactions while I am writing this review, for example. Of course they do. And, if we can discover them, they may well provide a complete explanation for the mechanisms of my brain while thinking and writing. The main claim of anti-reductionism in science is that such a complete explanation of mechanisms at one level does not necessarily explain what exists and happens at higher levels. Indeed we may need to know about the higher levels in order to explain the lower-level data that form an input to the mechanisms involved (which is what must be the case in writing this review! – one of the inputs was my reading this book, but the book is not thereby 'inside' my brain).

The most spectacular case of this need for higher-level understanding in modern biology is, of course, the genome, whose sequences will only be understood eventually in terms of higher-level function (genes don't come with functional names attached – nor do neurones!).

I started my life in physiological research as a fully paid-up member of the reductionist club. In the 1960s you couldn't get much more reductionist than to discover ion channels in excitable cells and then to simulate their activity in a bottom-up approach. I did for heart cells what Hodgkin & Huxley did for nerve cells. It is through trying to extend this approach to higher physiological levels that I have come to see the conceptual and computational problems that arise in practice. I have written elsewhere on the impossibility of a completely bottom-up reconstruction of living systems.[2]

Concurrently I also interacted extensively with professional philosophers (of different schools of thought – including those who would support the reductionist agenda). Coming to adopt an integrationist agenda was not an easy road, either scientifically or philosophically. But it is a far richer position. The integrationist does not deny the validity or immense achievements of successful reduction. For some reason (political, social, philosophical?) reductionists seem to need to claim universality for their approach. This book will give them some cause to re-think that position – or so I would hope.

Denis Noble CBE FRS hon FRCP
Professor of Cardiovascular Physiology at Oxford University
Secretary-General, International Union of Physiological Sciences (1993–2001)

---

[2] D. Noble, 'Biological computation', in *Encyclopedia of Life Sciences*, http://www.els.net (Nature Publishing Group, London, 2002); D. Noble, 'The rise of computational biology', *Nature Reviews Molecular Cell Biology*, 3 (2002), pp. 460–3. See also Novartis Foundation, *In Silico Simulation of Biological Processes*, Novartis Foundation Symposium, vol. 247 (John Wiley, London, 2002).

Acknowledgements to the Second Edition

In the course of writing this expanded second edition, we have benefited greatly from the advice and constructive suggestions of many friends and colleagues. We are grateful to Hanoch Ben-Yami, Anthony Kenny, Parashkev Nachev, Hans Oberdieck, Juan Pascual, Herman Philipse and Gabriele Taylor. They all read one or more of the new chapters or sections and gave us their helpful comments for which we are much indebted.

M. R. B.
P. M. S. H.

Acknowledgements to the First Edition

Many people have assisted us in writing this book, and we have benefited from their advice, critical comments and constructive suggestions. We are grateful to Dr Hanoch Benyami, Professor Jonathan Dancy, Professor John Dupré, Dr Hanjo Glock, Professor Oswald Hanfling, Professor Paul Harris, Dr Tim Horder, Professor Andrew Parker, Professor Herman Philipse, Dr John Richardson, Wolfram Schmitt and Professor Stuart Shanker for their helpful comments on one or more (sometimes many more) chapters of the book. We are especially indebted to Dr John Hyman and Professor Hans Oberdiek, who read and commented constructively on the whole text. We thank Sir Anthony Kenny, Professor Sir Peter Strawson and Professor Georg Henrik von Wright for their encouragement and moral support throughout the writing of this book. We are grateful to Jean van Altena for copy-editing our typescript with her characteristic skill, tact and good humour.

Chapters 1 and 2 are a rewritten version of a review article entitled 'The motor system in neuroscience: a history and analysis of conceptual developments', published in *Progress in Neurobiology*, 67 (2002), pp. 1–52. Parts of chapters 3, 4 and 5 were published in an article entitled 'Perception and memory in neuroscience: a conceptual analysis', in *Progress in Neurobiology*, 65 (2001), pp. 499–543. Parts of chapter 10 were published in an article by Hacker entitled 'Is there anything it is like to be a bat?' in *Philosophy*, 77 (2002), pp. 153– 70. We thank the editors of these journals for permission to republish this work.

M. R. Bennett, *University of Sydney*
P. M. S. Hacker, *St John's College, Oxford*

Introduction to the First Edition

Philosophical Foundations of Neuroscience presents the fruits of a cooperative project between a neuroscientist and a philosopher. It is concerned with the conceptual foundations of *cognitive* neuroscience – foundations constituted by the structural relationships among the psychological concepts involved in investigations into the neural underpinnings of human cognitive, affective and volitional capacities. Investigating logical relations among concepts is a philosophical task. Guiding that investigation down pathways that will illuminate brain research is a neuroscientific one. Hence our joint venture.

If we are to understand the neural structures and dynamics that make perception, thought, memory, emotion and intentional behaviour possible, clarity about these concepts and categories is essential. Both authors, coming to this investigation from very different directions, found themselves puzzled by, and sometimes uneasy with, the use of psychological concepts in contemporary neuroscience. The puzzlement was often over what might be meant by a given neuroscientist's claims concerning the brain and the mind, or over why a neuroscientist thought that the experiments he had undertaken illuminated the psychological capacity being studied, or over the conceptual presuppositions of the questions asked. The unease was produced by a suspicion that in some cases concepts were misconstrued, or misapplied, or stretched beyond their defining conditions of application. And the more we probed, the more convinced we became that, despite the impressive advances in cognitive neuroscience, not all was well with the general theorizing.

Empirical questions about the nervous system are the province of neuroscience. It is its business to establish matters of fact concerning neural structures and operations. It is the task of *cognitive* neuroscience to explain the neural conditions that make perceptual, cognitive, cogitative, affective and volitional functions possible. Such explanatory theories are confirmed or infirmed by experimental investigations. By contrast, conceptual questions (concerning, for example, the concepts of mind or memory, thought or imagination), the description of the logical relations between concepts (such as between the concepts of perception and sensation, or the concepts of consciousness and self-consciousness), and the examination of the structural relationships between distinct conceptual fields (such as between the psychological and the neural, or the mental and the behavioural) are the proper province of philosophy.

Conceptual questions antecede matters of truth and falsehood. They are questions concerning our *forms of representation*, not questions concerning the truth or falsehood of

empirical statements. These forms are presupposed by true (*and* false) scientific statements and by correct (*and* incorrect) scientific theories. They determine not what is empirically true or false, but rather what does and what does not make sense. Hence conceptual questions are not amenable to scientific investigation and experimentation or to scientific theorizing. For the concepts and conceptual relationships in question are *presupposed* by any such investigations and theorizings. Our concern here is not with trade union demarcation lines, but with distinctions between logically different kinds of intellectual inquiry. (Methodological objections to these distinctions are examined in chapter 14 [chapter 17 in this edition].)

Distinguishing conceptual questions from empirical ones is of the first importance. When a conceptual question is confused with a scientific one, it is bound to appear singularly refractory. It seems in such cases as if science should be able to discover the truth of the matter under investigation by theory and experiment – yet it persistently fails to do so. That is not surprising, since conceptual questions are no more amenable to empirical methods of investigation than problems in pure mathematics are solvable by the methods of physics. Furthermore, when empirical problems are addressed without adequate conceptual clarity, misconceived questions are bound to be raised, and misdirected research is likely to ensue. For any unclarity regarding the relevant concepts will be reflected in corresponding unclarity in the questions, and hence in the design of experiments intended to answer them. And any incoherence in the grasp of the relevant conceptual structure is likely to be manifest in incoherences in the interpretation of the results of experiments.

Cognitive neuroscience operates across the boundary between two fields, neurophysiology and psychology, the respective concepts of which are categorially dissimilar. The logical or conceptual relations between the physiological and the psychological are problematic. Numerous psychological concepts and categories of concepts are difficult to bring into sharp focus. The relations between the mind and the brain, and between the psychological and the behavioural, are bewildering. Puzzlement concerning these concepts and their articulations, and concerning these apparent 'domains' and their relations, has characterized neurophysiology since its inception (we shall begin our investigations in chapter 1 with a historical survey of the early development of neuroscience). In spite of the great advances in neuroscience at the beginning of the twentieth century at the hands of Charles Sherrington, the battery of conceptual questions popularly known as the mind–body or mind–brain problem remained as intractable as ever – as is evident in the flawed Cartesian views embraced by Sherrington and by such of his colleagues and protégés as Edgar Adrian, John Eccles and Wilder Penfield. Brilliant though their work unquestionably was, deep conceptual confusions remained – as we show in chapter 2. Whether the current generation of neuroscientists has successfully overcome the conceptual confusions of earlier generations, or whether it has merely replaced one conceptual entanglement by others, is the subject of our investigation in this book.

One such tangle is evident in the persistent ascription of psychological attributes to the brain. For, while Sherrington and his protégés ascribed psychological attributes to the mind (conceived as a peculiar, perhaps immaterial, substance distinct from the brain), contemporary neuroscientists tend to ascribe the same range of psychological attributes to the brain (commonly, although not uniformly, conceived to be identical with the mind). But the mind, we argue (§3.10), is neither a substance distinct from the brain nor a substance identical with the brain. And we demonstrate that ascription of psychological attributes to the brain is incoherent (chapter 3). Human beings possess a wide range of psychological powers,

which are exercised in the circumstances of life, when we perceive, think and reason, feel emotions, want things, form plans and make decisions. The possession and exercise of such powers define us as the kinds of animals we are. We may enquire into the neural conditions and concomitants for their possession and exercise. This is the task of neuroscience, which is discovering more and more about them. But its discoveries in no way affect the conceptual truth that these powers and their exercise in perception, thought and feeling *are attributes of human beings*, not of their parts – in particular, *not of their brains*. A human being is a psychophysical unity, an animal that can perceive, act intentionally, reason and feel emotions, a language-using animal that is not merely conscious, but also self-conscious – not a brain embedded in the skull of a body. Sherrington, Eccles and Penfield conceived of human beings as animals in whom the mind, which they thought of as the bearer of psychological attributes, is in liaison with the brain. It is no advance over that misconception to suppose that the brain is a bearer of psychological attributes.

Talk of the brain's perceiving, thinking, guessing or believing, or of one hemisphere of the brain's knowing things of which the other hemisphere is ignorant, is widespread among contemporary neuroscientists. This is sometimes defended as being no more than a trivial *façon de parler*. But that is quite mistaken. For the characteristic form of explanation in contemporary cognitive neuroscience consists in ascribing psychological attributes to the brain and its parts *in order to explain* the possession of psychological attributes and the exercise (and deficiencies in the exercise) of cognitive powers by human beings.

The ascription of psychological – in particular, cognitive and cogitative – attributes to the brain is, we show, also a source of much further confusion. Neuroscience can investigate the neural conditions and concomitants of the acquisition, possession and exercise of sentient powers by animals. It can discover the neural preconditions for the possibility of the exercise of distinctively human powers of thought and reasoning, of articulate memory and imagination, of emotion and volition. This it can do by patient inductive correlation between neural phenomena and the possession and exercise of psychological powers, and between neural damage and deficiencies in normal mental functions. What it *cannot* do is *replace* the wide range of ordinary psychological explanations of human activities in terms of reasons, intentions, purposes, goals, values, rules and conventions by neurological explanations (reductionism is discussed in chapter 13 [16]). And it *cannot* explain how an animal perceives or thinks by reference to the brain's, or some part of the brain's, perceiving or thinking. For it makes no sense to ascribe such psychological attributes to anything less than the animal as a whole. It is the animal that perceives, not parts of its brain, and it is human beings who think and reason, not their brains. The brain and its activities *make it possible* for *us* – not for *it* – to perceive and think, to feel emotions, and to form and pursue projects.

While the initial response of many neuroscientists to the accusation of conceptual confusion is to claim that the ascription of psychological predicates to the brain is a mere *façon de parler*, their reaction to the demonstrable fact that their explanatory theories *nontrivially* ascribe psychological powers to the brain is sometimes to suggest that this error is unavoidable due to the deficiencies of language. We confront this misconception in chapter 14, where we show that the great discoveries of neuroscience *do not require* this misconceived form of explanation – that what has been discovered can readily be described and explained in our existing language. We demonstrate this by reference to the much discussed phenomena resultant upon commissurotomy, described (or, we suggest, misdescribed) by Sperry, Gazzaniga and others (§14.3 [§17.3]).

In Part II we investigate the use of concepts of perception, memory, mental imagery, emotion and volition in current neuroscientific theorizing. From case to case we show that conceptual unclarity – failure to give adequate attention to the relevant conceptual structures – has often been the source of theoretical error and the grounds for misguided inferences. It is an error, a *conceptual* error, to suppose that perception is a matter of apprehending an *image* in the mind (Crick, Damasio, Edelman), or the production of a *hypothesis* (Helmholtz, Gregory), or the generation of a *3-D model description* (Marr). It is confused – a *conceptual* confusion – to formulate the binding problem as the problem of combining data of shape, colour and motion to form *the image* of the object perceived (Crick, Kandel, Wurtz). It is wrong, *conceptually* wrong, to suppose that memory is always of the past, or to think that memories can be *stored* in the brain in the form of the strength of synaptic connections (Kandel, Squire, Bennett). And it is mistaken, *conceptually* mistaken, to suppose that the investigation of thirst, hunger and lust is an investigation into the emotions (Rolls) or to think that the function of the emotions is to inform us of our visceral and musculoskeletal state (Damasio).

The initial reaction to such critical remarks may well be indignation and incredulity. How can a flourishing science be fundamentally in error? How could there be unavoidable conceptual confusion in a well-established science? Surely, if there are problematic concepts, they can easily be replaced by others that are unproblematic and that serve the same explanatory purposes. Such responses betoken a poor understanding of the relation between form of representation and facts represented, and a misunderstanding of the nature of conceptual error. They also betray ignorance of the history of science in general, and of neuroscience in particular.

Science is no more immune to conceptual error and confusion than any other form of intellectual endeavour. The history of science is littered with the debris of theories that were not simply factually mistaken, but conceptually awry. Stahl's theory of combustion, for example, was *conceptually* flawed in ascribing, in certain circumstances, negative weight to phlogiston – an idea that made no sense within its framework of Newtonian physics. Einstein's famous criticisms of the theory of electromagnetic aether (the alleged medium by which light was thought to be propagated) were directed not only at the results of the Michelson–Morley experiment, which had failed to detect any effect of absolute motion, but also at a conceptual confusion concerning relative motion involved in the role ascribed to aether in the explanation of electromagnetic induction. Neuroscience has been no exception – as we show in our historical survey. It is true enough that the subject is now a *flourishing* science. But that does not render it immune to conceptual confusions and entanglements. Newtonian kinematics was a flourishing science, but that did not stop Newton from becoming entangled in conceptual confusions over the intelligibility of action at a distance, or from bafflement (not remedied until Hertz) over the nature of force. So too, Sherrington's towering achievement in explaining the integrative action of synapses in the spinal cord, and thereby eliminating, once and for all, the confused idea of a 'spinal soul', was perfectly compatible with conceptual confusions concerning the 'cerebral soul' or mind and its relation to the brain. Similarly, Penfield's extraordinary achievements in identifying functional localization in the cortex, as well as in developing brilliant neurosurgical techniques, were perfectly compatible with extensive confusions about the relation between the mind and the brain and about the 'highest brain function' (an idea borrowed from Hughlings Jackson).

In short, conceptual entanglement *can* coexist with flourishing science. This may appear puzzling. If the science can flourish despite such conceptual confusions, why should scientists care about them? Hidden reefs do not imply that the seas are not navigable, only that they are dangerous. The moot question is how running on these reefs is manifest. Conceptual confusions may be exhibited in different ways and at different points in the investigation. In some cases, the conceptual unclarity may affect neither the cogency of the questions nor the fruitfulness of the experiments, but only the understanding of the results of the experiments and their theoretical implications. So, for example, Newton embarked on the *Optics* in quest of insight into the character of colour. The research was a permanent contribution to science. But his conclusion that 'colours are sensations in the sensorium' demonstrates failure to achieve the kind of understanding he craved. For, whatever colours are, they are not 'sensations in the sensorium'. So in so far as Newton cared about *understanding* the results of his research, then he had good reason for caring about the conceptual confusions under which he laboured – for they stood in the way of an adequate understanding.

In other cases, however, the conceptual confusion does not so happily bracket the empirical research. Misguided questions may well render research futile (examples will be examined in relation to mental imagery (§6.3.1 [§7.3.1]) and voluntary movement (§8.2 [§9.2])). Rather differently, misconstrual of concepts and conceptual structures will sometimes produce research that is by no means futile, but that fails to show what it was designed to show (examples will be discussed in relation to memory (§§5.2.1–5.2.2 [§§6.2.1–6.2.2]) and to emotions and appetites (§7.1 [§8.1])). In such cases, the science may not be flourishing quite as much as it appears to be. It requires conceptual investigation to locate the problems and to eliminate them.

Are these conceptual confusions *unavoidable*? Not at all. The whole point of writing this book is to show how to avoid them. But, of course, they cannot be avoided while leaving everything else intact. They *can be* avoided – but if they are, then certain kinds of questions will no longer be asked, since they will be recognized as resting on a misunderstanding. As Hertz put it in the wonderful introduction to his *Principles of Mechanics*: 'When these painful contradictions are removed, … our minds, no longer vexed, will cease to ask illegitimate questions.' Equally, certain kinds of inferences will no longer be drawn from a given body of empirical research, since it will be realized to have little or no bearing on the matter which it was meant to illuminate, even though it may bear on *something else*.

If there are problematic concepts, can they not be replaced by others that serve the same explanatory function? A scientist is always free to introduce new concepts if he finds existing ones inadequate or insufficiently refined. But our concern in this book is not with the use of new technical concepts. We are concerned with the misuse of old, nontechnical concepts – concepts of mind and body, thought and imagination, sensation and perception, knowledge and memory, voluntary movement, and consciousness and self-consciousness. There is nothing inadequate about these concepts relative to the purposes they serve. There is no reason for thinking that they need to be replaced in the contexts that are of concern to us. What are problematic are neuroscientists' misconstruals of them and the misunderstandings consequently engendered. These are remediable by a correct account of the logico-grammatical character of the concepts in question. And this is what we have tried to supply.

Granted that neuroscientists may not be using these common or garden concepts the way the man in the street does, with what right can philosophy claim to correct them? How