

Neuroscience Through the Looking Glass

M.R. Bennett & P.M.S. Hacker. (2003). *Philosophical foundations of neuroscience*. Oxford, UK: Blackwell. ISBN 1-4051-0855-X (hardcover), 1-4051-0838-X (pbk). 461 pp.

Neuroscientist's wife: John, where did you see Mary last time?

Neuroscientist: In my occipital cortex, of course.

This book is a strong dissonance in a loud chorus of neuroscientists, psychologists, and philosophers who promise that, after yet another few years, one more large-scale project, one more combined effort of EEG, MEG, fMRI, and PET experts, one more imaging technology, the mystery of mind and consciousness, of feelings and thoughts, will finally be solved (or dissolved). Then the dualism of mind and matter will be overcome and replaced by a long-awaited unified science of mind/brain/gene, which will explain all the heights and subtleties of the human condition on the basis of neuronal activity – or at the very least, the chemical and physical processes in the brain. In contrast to this view (very optimistic or frightening, depending on one's perspective), Bennett and Hacker (hereafter referred to as B&H) suggest that modern neuroscience – perhaps not the whole of it, but at least large parts – is now at a dead end, not because it did not develop technologies to answer more and more intriguing questions, but because it cannot formulate the correct questions. Not because it is unable to reach its goals, but because it has no idea which goals should be reached. Not because of the unsolved experimental problems, but because of the conceptual ones.

Those who take up the task of reading the 450 pages of this book can accept or reject one or more of the ideas and conclusions of the authors – but they will be unable to go back to “science as usual,” passively accepting the dogmas so strongly shaken by B&H. This means that the book is mandatory reading material.

The book has several important merits. First, the tandem of prominent authors – a neurophysiologist and a philosopher – are fully devoted to the search for truth and clarity, not to impress the public with a new bestseller. Second, it is written in excellent English, with a brilliant sense of humor. Third, it is extremely comprehensive and in this respect, reminds one of James's *Psychology*. The authors consecutively analyze all forms of mental activity: sensation, pain, perception, imagination, memory, thinking, knowledge and beliefs, emotions, volition, and different kinds and subdivisions of consciousness. In addition to this, B&H briefly review the history of philosophical ideas as manifested in the classics of neuroscience from Willis and Bell to Adrian, Eccles, and Pen-

field. They discuss methodological issues in the study of the mind-body problem and formulate a clear, though not uncontroversial, view on what is possible and impossible for neuroscience, as well as what is possible and impossible for philosophy. They close with an analysis of the several most influential approaches in the modern philosophy of mind, as represented in the recent books of the Churchlands, Dennett, and Searle. Thus, the authors do not leave out any of the important problems of interest for cognitive and behavioral neuroscience.

Their main conclusions can be summarized as follows: (1) Most modern neuroscience is based on a confused and unnatural mixture of Cartesianism, traditional British empiricism (Locke, Berkeley), and materialism. Rejecting the mind/body dualism of Descartes, modern neuroscientists simply adopted (and slightly adapted) everything else in his doctrine, simply re-ascribing all the presumable functions of the mind – to the brain (or even, following Chomsky, to the hybrid of mind-brain). From Locke and Berkeley, neuroscience took on subjectivism, as well as the notion that in perception, imagery, and memory the mind-brain creates “images” or “copies” of external objects – good copies in perception, weaker in imagery, yet weaker in memory. (2) As a result of this philosophical confusion, a big part of neuroscience is busy with continuous attempts to solve nonexistent problems, i.e., problems that emerge solely because of the erroneous use of words, such as introducing metaphors that are then conceived of literally. Examples of such metaphors are “maps” (as if there were open books in the brain for orientation in space) or “representations” (the verb “to represent,” like “to give,” requires a dative: to whom?).

The paradigmatic case of such a pseudo-problem is the so-called binding problem, which only exists for one who believes that what we see are not real objects in the world, but small pictures of these objects in the brain. Of course, such pictures cannot exist, because if they exist, the homunculus who observes them would have to have his own mechanisms of visual perception in order to recognize the images the visual cortex shows him, and so on, ad infinitum. But as soon as we understand that “images in the brain” is merely a metaphor, that in order to see an apple we need neither a realistic painting nor an abstract drawing of it in the occipital cortex, the fact that some neurons selectively respond to some features rather than to other features, does not present a problem. B&H point out that the first version of the “binding” problem was invented by Descartes himself. He argued that the structure in which the brain contacts the immortal soul must lie exactly on the midline because, if it is lateralized, the soul would receive

more excitation from one side of the body than from the other side, and, hence, we would perceive the world skewed to one side. Are you laughing? But why are you not laughing when a modern scientist maintains that there must be a specific mechanism uniting all discharges of specialized neurons, otherwise we would perceive the world split into single lines, angles, color patches, etc.? Such a mechanism is no more necessary than a specific mechanism to turn the inverted retinal image upright again. Note that B&H regard highly the discovery of Singer and Gray that synchronous firing of the neurons is an important condition for coherent vision (Singer, 1999). This does not, however, have anything in common with the problem of feature binding (Treisman & Gelade, 1980), because there is nothing to bind. Sometimes metaphors can be innocuous but they are traps that neuroscientists so frequently fall into.

Another example is the famous experiment of Libet, Gleason, Wright, & Pearl (1983), in which subjects were asked to make self-paced simple finger movements and to report, after each movement, at what time point they had first experienced an "urge to move" their finger. This was regarded as an experimental model of a "freely voluntary action," and the data of this experiment became a basis for numerous discussions on physiological foundations of free will or the lack thereof (Haggard & Eimer, 1999; Gomes, 2002; Klein, 2002; Rosenthal, 2002; Libet, 2003). However, as B&H indicate, this is completely misleading because there is no such thing as the "urge" to any activity we call voluntary. We can feel an urge to sneeze, or cough, or scratch some area on the skin, all these being rather typical examples of *involuntary* movements, even though they are under partial voluntary control in so far as, if necessary, we can (try to) suppress them. Voluntary actions are locomotion, or applying for a new job, or picking up a glass of water. None of these actions involves any introspection of feelings whose timing Libet's subjects had to judge. Some rather complex behaviors often involve an experience of having made a decision, but the decision is based on reasons, not on an internal "wish," let alone "urge," to carry out the corresponding action. Also, we sometimes feel a "desire," but this is neither necessary nor sufficient for an act to be voluntary. The logical confusion into which both Libet and his critics fell is demonstrated by B&H with great clarity, on just two pages (229–230). Whoever reads these pages once will lose interest in this absurd discussion forever.

B&H propose simple and lucid descriptions of such "mysterious" phenomena as blindsight and behavioral consequences of commissurotomy (the description can easily be extended to other dissociation conditions, e.g., neglect), and show that, when appropriately depicted, these phenomena contain no mystery at all. Normally, we use two legs for locomotion, and their movements are

coordinated. After a typical infarct in a capsula interna, movements in the contralateral leg are severely impaired, but movements in the ipsilateral leg remain intact. Thus, two functions that are normally closely associated become dissociated after a particular brain lesion. This paresis is an important phenomenon, but nobody finds it mysterious. Likewise, functions of pointing to a visually presented object and naming it, closely associated in healthy individuals, can be dissociated due to a brain lesion. But this dissociation only appears mysterious when described in obscure terms.

In such a large and all-embracing book, not everything is equally persuading. For me, dozens of questions raised by B&H remained open, and the answers suggested by the authors were not convincing. As I am writing a book review, and not another book, I will restrict myself to a few topics. First, B&H explicitly follow the Wittgensteinian approach to science and philosophy. (Hacker is perhaps the most prominent expert on Wittgenstein today.) Wittgenstein's thinking passed through several profound changes during his life, but he focused on a number of ideas that are strongly apparent in this book. One of them is the strict distinction between conceptual problems on the one hand, and empirical problems on the other. The former are philosophical problems that can be solved by means of linguistic analysis, which allows us to distinguish between meaningful and meaningless statements. The latter are scientific problems that can be solved by empirical studies, which allow us to distinguish between true and false statements (both of them being meaningful). Thus, philosophy is not about the World, but only about the Word. Its task is clarification of scientific propositions, not generation of its own propositions (Tractate, 4.112; cf. Wittgenstein, 1963). This radical distinction between philosophical and scientific problems was widely accepted in the period between the two World Wars, but it was strongly criticized in the second half of the 20th century, particularly, by Popper and his followers, who argued that problems which appear conceptual within one conceptual system may become empirical in another system (Popper, 1963). In fact, when B&H try to explain why this or that statement is a conceptual truth, they frequently indicate that this is because that is how we learn to use the corresponding words. But this is an argument from empirical science: the theory of socialization and developmental psychology.

A consequence of this approach is the belief that everything that we are, we are because of our language. "*The limits of my language mean the limits of my world*" (Tractate, 5.6; italics in original). What cannot be described in words (or other symbols: diagrams, mathematical formulas, musical notes, etc.) does not exist for us. Not everybody would agree with this; Buddhists surely not, but they're not alone. To understand a concept (e.g., "red"), B&H state, does not mean to relate it to any experience,

but to be able to use the corresponding word appropriately. "Red" is, for example, by definition darker than pink and closer to orange than to yellow. But this is simply wrong. Kay & Kempton (1984) presented their subjects with two green colors and one blue. One of the green colors (Green 1) was separated from the other (Green 2) by a larger number of just noticeable differences than from Blue. Surely, all subjects correctly named Green green, and Blue blue. However, when asked to compare stimuli pairwise, they found Green 1 and Green 2 *less alike* than Green 1 and Blue. Thus, contrary to B&H, not only need red (more exactly, a case of red) not be closer to orange than to yellow, red can be closer to orange than to red! When we compare colors, we do not compare definitions of the respective words (Kotchoubey, 2005b).

This line of reasoning leads B&H to almost complete negation of the private aspects of the human condition. They discuss a possible counterargument that our primary experiences are incommunicable and indescribable in words – indeed, "how to describe an experience of smelling freshly ground coffee, of hearing an oboe being played, or of seeing the Mediterranean summer sky?" Their answer is, very easy: "The first is very pleasant . . . ; the second may be absolutely wonderful . . . ; and the last may be intoxicating and enthralling" (p. 286). Such platitudes! In vain I looked for a sign of irony around these words. The authors appear to believe seriously that a person who had never heard an oboe would get at least a slight impression of it by being told that it is "wonderful."

Thus, according to B&H, mental states (sensations, feelings, etc.) are not private but public, they are just products of our acquisition of the corresponding mental terms: verbs like "to see," "to believe," "to be angry," etc. Hence, we need not infer such states in other people (as well as higher animals) – rather, we immediately *see* them in a person's behavior, except when he/she makes a particular effort to conceal them. A man is in pain if he expresses pain-related behaviors. We don't suppose, or conjecture, or believe that another person is conscious, or in pain, or feels something – rather, as B&H repeat several times, "we know this perfectly well." One might notice that, despite B&H's hostility toward Cartesianism, the idea that a truth just reveals itself if we look at it free from conceptual confusions is a legacy of classical philosophical rationalism of the 17th century (perhaps more Spinozean than Cartesian).

However, from the empirical point of view, this Wittgensteinian stance has already been demonstrated to fail in patients with severe neurological disorders, especially the locked-in syndrome (Kurthen, Moskopp, Linke, & Reuter, 1991). B&H also implicitly admit this fact when they talk about "singularities." Were the blink reflex of Jean-Dominique Bauby paralyzed, nobody would have ever known what a beautiful mind was living in his totally immobile

body (Bauby, 1997). Ascription of (nontransitive) consciousness to severely disabled patients is a big problem, quite different from the sweet but wrong story (told on p. 247) of a patient who awakes from coma. Sometimes indirect physiological data are the only basis for judging a patient's mental state (Kotchoubey, Lang, Bostanov, & Birbaumer, 2003). To make a clinically and ethically responsible decision, we use a plethora of objective sources of information on the remaining brain and bodily functions, and then we *infer* that the observed level of these functions is *probably* indicative of the patient being conscious (Kotchoubey, 2005a). Disregarding such cases as pathological is irrelevant because the authors insist on the "conceptual link" – not an empirical correlation – between mind and behavior. What meaning has "pathology" in a conceptual truth? What would it mean to say that "triangles cannot be spherical," but that "some exceptional, severely ill triangles can"? It may be true that *usually* "we attribute consciousness to a creature on the grounds of its behavior in the circumstances of its life, not on the grounds of its possessing private qualia," but this is an empirical hypothesis in the domain of social psychology. It is not a philosophical statement at all and says nothing about the existence or nonexistence of qualia, like the statement "usually, most people think more on money than on God" is neither a proof nor a disproof of the existence of God.

Like myself, other readers will surely disagree with many of B&H's views. I do not think that any neuroscientist or psychologist reading the book would accept everything in it. One does not read such a book in order to agree with each line of it, but to look from another perspective onto one's own, and our common, research practice. The book is very timely. Two centuries after Kant, it is time also for neuroscience to "awake from the dogmatic sleep" and, regaining a scientific (meaning critical) self-consciousness, to become aware of the lot of nonsense we are researching and publishing.

References

- Bauby, J.-D. (1997). *Le scaphandre et le papillon* (1st ed.). Paris: Editions Robert Laffont.
- Gomes, G. (2002). Problems in the timing of conscious experience. *Consciousness and Cognition*, 11, 191–197.
- Haggard, P., & Eimer, M. (1999). On the relation between brain potentials and the awareness of voluntary movements. *Experimental Brain Research*, 126, 128–133.
- Kay, P., & Kempton, W. (1984). What is the Sapir-Whorf hypothesis? *American Anthropologist*, 86, 65–78.
- Klein, S. (2002). Libet's research on the timing of conscious intention to act: A commentary. *Consciousness and Cognition*, 11, 273–279.
- Kotchoubey, B. (2005a). Apallic syndrome is not apallic – is vegetative state vegetative? *Neurological Rehabilitation*, in press.

- Kotchoubey, B. (2005b). Words, colors, and robots: Whorf wouldn't be satisfied. *Behavioral and Brain Sciences*, in press.
- Kotchoubey, B., Lang, S., Bostanov, V., & Birbaumer, N. (2003). Cortical processing in Guillain-Barre syndrome after years of total immobility. *Journal of Neurology*, 250, 1121–1123.
- Kurthen, M., Moskopp, D., Linke, D.B., & Reuter, B.M. (1991). The locked-in syndrome and the behaviorist epistemology of the minds. *Theoretical Medicine*, 12, 69–79.
- Libet, B. (2003). Timing of conscious experience: Reply to the 2002 commentaries on Libet's findings. *Conscious and Cognition*, 12, 321–331.
- Libet, B., Gleason, C.A., Wright, E.W., & Pearl, D.K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). The unconscious initiation of a freely voluntary act. *Brain*, 106, 623–642.
- Popper, K.R. (1963). *Conjectures and refutations*. London: Routledge and Kegan Paul.
- Rosenthal, D.M. (2002). The timing of conscious states. *Consciousness and Cognition*, 11, 215–220.
- Singer, W. (1999). Neuronal synchrony: A versatile code for the definition of relations. *Neuron*, 24, 49–65.
- Treisman, A.M., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology*, 12, 97–136.
- Wittgenstein, L. (1963). *Tractatus logico-philosophicus* (17th ed.). Frankfurt: Suhrkamp.

Address for correspondence

Boris Kotchoubey
Department of Medical Psychology and Behavioral
Neurobiology
University of Tübingen
Gartenstr. 29
D-72074 Tübingen
Germany
Tel. +49 7071 29-74221
Fax +49 7071 29-5956
E-mail boris.kotchoubey@uni-tuebingen.de
