



## Book review essay

# Consciousness studies: The view from psychology

### **Consciousness: A user's guide**

By Adam Zeman

Yale, CT: Yale University Press. 2003. Hbk £18.95, ISBN 0-300-09280-6/2004. Pbk £10.99, ISBN 0-300-10497-9

### **Mirror neurons and the evolution of brain and language**

Edited by Maxim I. Stamenov and Vittorio Gallese

Amsterdam: John Benjamins. 2002. Hbk €110.00, ISBN 90-272-5166-5. Pbk €72.00, ISBN 90-272-5162-2

### **Consciousness: Creeping up on the hard problem**

By Jeffrey Gray

Oxford: Oxford University Press. 2004. Hbk £29.95, ISBN 0-19-852090-5. Pbk ISBN 978-019-852090-0

### **Consciousness: An introduction**

By Susan Blackmore

London: Hodder and Stoughton. 2003. Pbk £16.99, ISBN 0-3408-09094

In this review essay, I discuss four recent books ranging from philosophical to neuropsychological approaches to consciousness. The books are just a small selection from a large number of recent texts that have not only been insisting that we should take consciousness as a serious subject for scientific study, but which also have identified themselves with the new field of 'consciousness studies'. The books in review represent not just new publications of this line of study. They also represent a typical variety of approaches, mirroring different interpretations of 'the problem of consciousness'.

Consciousness studies has developed over the last 20 or so years as an autonomous, yet interdisciplinary, field of research and theory, embracing philosophy, psychology and neuroscience. Since consciousness studies have yet to become part of the curriculum of mainstream psychology, I first provide an outline of the field of consciousness studies, and its central themes and problems. Furthermore, as the books under review themselves demonstrate, there is no consensus about what even constitutes the problem of consciousness, although each takes consciousness as its central topic. Each of the books attempts, in its own way, to provide an overview of this new field of research, and so, having reviewed each according to its own merits, I make comparisons among them. Finally, since the books offer their own projections about the future of consciousness studies, I close this review with my own reflections concerning the prospects for this new science and its possible relations with psychology.

### The history of consciousness studies

Unravelling the history of the new interest in consciousness is surprisingly difficult, even though it has been around only for a few decades. One possible reason for this is the fact that the field of consciousness studies is defined interdisciplinarily. In philosophy, it is commonly held that the interest in consciousness, at least formulated as a 'mind-body problem', came about with René Descartes. Descartes famously claimed that subjective conscious states are separate from the world and that they causally interact with the world through their physical seat, the brain. While this framework has long been abandoned, philosophical debates about consciousness have nevertheless largely been focused on the metaphysics of consciousness. How can something essentially subjective – the phenomenal experiences of a given person or animal – exist in a material world? The debate has primarily involved the materialists, who claim that phenomenal consciousness, in spite of our everyday intuition, actually refers to physical phenomena (brain processes), on the one side, and the dualists on the other, who insist that consciousness is neither identical nor reducible to something material. In experimental psychology, however, the story is quite different. In the late nineteenth century, the German, American and French introspectionists set out to study consciousness experimentally by varying physical stimuli and collecting reports of the corresponding subjective variations. As a result of many methodological conflicts in which different laboratories kept failing to reproduce each others' results, introspectionism declined and the more experimentally promising paradigm of behaviourism took over as a leading trend in psychology. As behaviourism in its most radical versions famously denied the existence (or, as in Skinner's case, the explanatory relevance) of 'internal' or 'mental states', so did theorizing about – or even the very mention of – consciousness disappear from mainstream psychology.

Since the 1980s, and some 30 years after the 'cognitive revolution', there has been a growing revival of interest in consciousness among psychologists, philosophers and neuroscientists. As the history of the revival of interest in consciousness is different in the various disciplines contributing to its study, so the reasons for this revival are also diverse. Several recent philosophical articles and books have discussed paradigms within experimental psychology, especially the currently dominant paradigm of cognitive psychology. Their important criticism of cognitive science has been that it has not succeeded in moving on from behaviourism in its fundamental methodology and assumptions (Leahey, 1987). First and foremost, the argument has been that cognitivist theories are unable to embrace consciousness. Basically, the kinds of models that cognitive theories adopt – algorithmical modelling and 'boxes with arrows between them' – do not in any relevant way describe what our experiences are like to ourselves. Indeed, this kind of 'boxology' could *never* do so, as a matter of principle. Therefore, we need a different experimental framework to deal with *conscious* rather than objectified subjects. If not, it is argued, we will accept a psychological framework that is at best agnostic when it comes to the one feature that is most prominent and clearly present in our everyday lives – that is, consciousness. Within experimental psychology itself, an increasing number of scientists have found it useful to apply the concept of consciousness at both a descriptive and explanatory level to phenomena known to psychology. For well-known phenomena such as subliminal perception, it has proved reasonable to speak of being 'above' or 'below' a perceptual threshold as a matter of being either in a conscious or unconscious perceptual state. This, also, has a 'phenomenological validity' – that is, the concepts actually address what it is like for a subject to experience being above or below threshold. Other examples of findings that were easily described with an appeal to consciousness have been neuropsychological syndromes such as blindsight and agnosia, where patients prove to be informed about worldly happenings even though they are unconscious of them. The ambition quickly arose, based upon such contrasts between consciousness and unconsciousness, to use consciousness as an experimental variable in the search for neural correlates of consciousness.

By the beginning of the 1990s, enthusiasm was growing for the development of a new and separate, interdisciplinary 'science of consciousness'. This development was in part, at least, inspired by a paper by Francis Crick and Christof Koch entitled 'Toward a neurobiological theory of consciousness' (1990) and kept alive by the annual conference series *Toward a Science of Consciousness*, organized by the Center for Consciousness Studies at the University of Arizona.

Even though we here also find attempts to discuss ontological issues about consciousness, the genuinely new development is the very idea that there could be a distinctive *science* of consciousness. That is, the suggestion is that, with some degree of methodological innovation, cognitive science or neuroscience could *directly* target consciousness as an object of empirical study, and not simply, as had previously been the case, try to make claims about the nature of consciousness based upon experiments that had actually been developed to address quite different issues. Consciousness studies may thus be said to have arisen as a theoretically guided change of focus and explanation in empirical research. At least, the attempt to arrive at an empirical science of consciousness is thought by many to be the end goal of the theoretical work.

Regardless of the aspirations of those working with consciousness studies for a new science, some, like Bernard Baars (1988), have argued that we do *already* have a science of consciousness, namely, cognitive psychology. This is, however, a rather extreme view. There is wide agreement that there is a *methodological* version of the mind-body problem, and that this problem arises precisely because the current methods of cognitive science fail to address consciousness. In order to counter the rather complacent view expressed by Baars, it is necessary to distinguish between two research paradigms within the overall discipline called cognitive science. I shall call them 'classical cognitive science' and 'cognitive neuroscience'. In some respects, these two approaches are similar: classical cognitive scientists and cognitive neuroscientists would certainly share some interest in the same experimental research (although the cognitive neuroscientists might be less interested in experiments having nothing to do directly with the brain), and they would be in perfect agreement about how to analyse data. However, on closer inspection, they prove to be linked to two very different theoretical backgrounds.

The cognitive science approach was the first to develop. George Miller attributes its 'consensual birthdate' to 11 September 1956, when a symposium on information theory was held at MIT (Miller, 1979, 2003). Cognitive science developed its methodology and theory through such classical works as Noam Chomsky (1957), George Miller (1956), George Sperling (1960) and Allen Newell (Newell & Simon, 1972). The tradition is still very much alive today and part of the standard curriculum of psychology. What unifies this tradition is the commitment to the algorithmic level of description for identifying the functional procedures that must be carried out in order to produce a behavioural response and thus give us all the necessary insight into the relation between mental and (other) physical phenomena. This is theoretically very close, for example, to Hilary Putnam's (1967) philosophical tradition of machine functionalism. The central analogy is that the mind is to the brain as a computer program is to the hardware of the computer. States of the program can be specified in a mathematical form and the physical state playing the role of carrying out those states will be irrelevant or at least extremely loosely connected to the state itself. Moreover, since the 'mind program' could be realised in different kinds of substrate (known as the principle of multiple realizations), we could transfer knowledge about the mind-brain relation to create a mind-computer relation. However, even if we accept this software-hardware analogy, there are problems. For instance, John Searle has introduced the 'Chinese Room argument' as an argument against machine functionalism (Searle, 1980, 1992). Say we were to place an English-speaker in a closed room and gave him certain (highly complicated) rules to follow - when you are presented with these signs, answer in this way. Given the instructions and the learning capabilities of the subject are sufficiently advanced, it will be possible to get the English-speaker to simulate a Chinese-speaker perfectly through writing without ever understanding to what he was actually responding. Or - as Searle concludes - semantics cannot be reduced to syntax. Put in another way, acting out a programme is not identical to understanding or knowing. Similarly, Ned Block has introduced the 'Chinese Giant argument'. Suppose we get the population of China to act out the behaviour of a set of neurons, running the right 'programme of mind'; then we would have a giant thinker indeed. With such examples, the sole dependency of mind on function seems to result in intuitively absurd conclusions (Block, 1978).

A different line of criticism against functionalism argues that the relation between cognition and consciousness may be strictly contingent (Chalmers, 1996). David Chalmers asks us to imagine a creature that atom by atom is identical to a human being, but who lacks qualia. Chalmers

argues that this is logically conceivable (although probably not factual) since no logical rule enables us to deduce an organism's subjective states from its behaviour: although we can collect reports from people claiming to be conscious when they are in certain physiological states, we cannot conclude that whenever this physiological state is present, a subject will necessarily be conscious. The contingent relation between consciousness and brain states does not rule out exceptions – that a 'zombie' would issue the report that it is in fact conscious when entertaining the same brain states that normal subjects have when issuing identical reports. The fundamental criticism against cognitive science is, then, that it is inherently incapable of telling zombies from normal subjects or, in other words, cognitive science studies all of its 'subjects' as if they were zombies.

There is, however, as I have already noted, another theoretical paradigm within cognitive science that is sometimes referred to as cognitive neuroscience. Cognitive science has proved itself methodologically compatible with neuroscience, which, in turn, has opened up the possibility of a cognitive neuroscience. Techniques such as PET, fMRI and EEG have made it possible – along with the older methods of studying brain-damaged patients and animals with induced lesions – to determine fairly precise neural correlates of cognitive functions. In this manner, cognitive neuroscience is dealing with two levels of description, and yet neither of these levels in any *a priori* way entails or explains the other.

Quite often, cognitive neuroscience seems to be based on some variety of identity theory, although there are some important exceptions among the brain scientists, such as John Eccles, who was committed to Cartesian dualism. Although the methodologies of cognitive science and cognitive neuroscience thus do not necessarily entail a commitment to monistic positions as their paradigms of thought, counter-examples are rare. On the contrary, it is much more often argued that these varieties of empirical psychology can solve the mind-body problem (Baars, 1988). Based on a largely implicit commitment to identity theory, consciousness is equated with a brain process or a function by cognitivists. The strategy has been to construct an information-processing model to account for some aspects of the brain's activity and then to 'explain' consciousness by identifying it with some part of the model (see Shallice, 1972, for one of the earliest examples of this).

In this manner, consciousness gets confused with its geographical location, and behavioural data – which are indeed relevant to the functions with which the brain is theoretically associated – come, however, to be regarded as information about consciousness. But, in fact, consciousness has nothing at all to do with the workings of a cognitive model such as Shallice's (1972): take away conscious experience and the model continues to work as before. In, say, computer simulations, consciousness might, in principle, be identical to some function, but we could never test for the occurrence of consciousness in the computer because there are no output events for which it could be said that consciousness is a necessary precursor. This point is more or less identical to David Chalmers' zombie argument, but was put forward much earlier by Peter White (1980).

Given we still lack a coherent scientific framework for empirical consciousness studies, it remains an open question which paradigmatic turn might be taken by consciousness research in the future. More specifically, it is an open question whether consciousness will become even more detached from cognitive science and neuroscience, or more integrated with those disciplines.

Consciousness studies has now been around for more than two decades as a line of thought and research separate from cognitive science. George Miller, one of the pioneers of cognitive science, long ago complained that 'consciousness is a word worn smooth by a million tongues. Depending upon the figure of speech chosen it is a state of being, a substance, a process, a place, an epiphenomenon, an emergent aspect of matter, or the only true reality' (Miller, 1962, p. 25). Today, 42 years later and after the enormous rise of interest in consciousness, a prominent author of cognitive theories about consciousness has come to a similar conclusion: 'Nothing has been resolved. A vast amount of brain power has been devoted to more and more minute examinations of questions that are not one step closer to solution today than in 600 BCE' (Baars, 2003, p. 32). To me, it seems almost bizarre that a research field receiving so much interest has made so little progress in so many years (at least, any 'progress' that all the researchers involved would regard as such).

If one considers the state of the art in consciousness studies, the lack of unity is very apparent. The terms 'the problem of consciousness' and 'the mind-body problem' are used as though they have a clear meaning. Yet, not only do we encounter many different interpretations of the nature of consciousness but, as Miller points out, the very framing of the problem is also subject to disagreement. The very concept of consciousness has several different meanings and it is often uncritically applied to mean different things. It seems almost unavoidable that a definition of consciousness will be circular (saying, for example, that consciousness is 'phenomenality' or 'experience', which then in turn can be defined as consciousness). Are we to arrive at one single definition of consciousness or rather is the field concerned with a whole set of different phenomena? See, for example, Block (1995), Allport (1988) and Nelkin (1993). Thus, although there is a growing consensus that consciousness poses an important scientific 'problem', this problem has come to be formulated in a number of distinctly different ways, as the books under review themselves explicitly address or implicitly demonstrate.

Before turning to the books themselves, I first identify the main current formulations of the nature of consciousness and the different definitions of the problem it is supposed to pose.

- (1) *The epistemological interpretation.* Our knowledge of objects in the world is 'third-person knowledge'. We experience objects from the outside, and our knowledge about them is not identical to the observed objects themselves. Our knowledge of our own experiences, however, is 'first-person knowledge'. The subject having the experience has a direct acquaintance with it, whereas other subjects can only be informed about it by way of reports issued by the experiencing subject. Thus, when we examine the brain, we do not observe experiences in there - only greyish-white tissue. On the other hand, when we introspectively examine our own experiences, we have no feeling of brain processes being their cause. All in all, therefore, knowledge about and the 'perceiving of' our own conscious experiences must be regarded as a completely different matter from our knowledge of external objects (see, e.g. Jackson, 1986; McGinn, 1989, 1991; Nagel, 1974, 1986).
- (2) *The ontological interpretation.* What is the 'nature' of consciousness? Is consciousness essentially different from physical matter or is it nothing but physical matter? We can observe close correlations between psychological, behavioural, cognitive and neurobiological phenomena, but such correlations do not imply specific causal relations between those phenomena. The problem of consciousness, according to this view, is how to understand consciousness in relation (e.g. causal relation) to other things. This interpretation is possibly the most widespread one, and it is debated in, for example, Churchland (1986), Chalmers (1996), Searle (1992), Kripke (1980) and Rosenthal (1990).
- (3) *The methodological interpretation.* How can we scientifically study something that is subjectively defined? Obviously, we cannot apply standard scientific methods to subjective data. Do we need a new 'science of consciousness' or can we, with few adjustments, still rely on contemporary cognitive psychology or neuroscience? This issue has been discussed by, for example, Chrisley (2001, 2002), Nagel (1986) and Petitot, Varela, Pachoud, and Roy (1999).

Each of the above represents different interpretations of the alleged problem of consciousness. It is, of course, not a matter of opting for the correct interpretation - rather, the problem of consciousness is indeed a problem in every one of the abovementioned ways according to most theorists' account. The attitude expressed in Baars (2003) (see above) may fit the status of the ontological interpretation of the problem of consciousness (although many distinguished philosophers would disagree), but has nothing to do with, say, the methodological or definitional interpretation. Thus, one could agree, although many would not, that little has happened by way of new developments in the clarification of the ontological or epistemological issues related to the mind-body problem. However, the methodologically focused debate about consciousness in relation to, say, neuroscience, is left untouched by any claim of stagnation. Still, of course, it is highly motivated by ontological questions.

Although philosophers have been debating consciousness for centuries, there has been much new controversy over the topic during the last 100 years, and the theoretical positions have varied tremendously. Consciousness has been conceived of as 'the starting point of psychology' (James, 1890), but also, in contrast, as 'uninfluential' and even 'non-existing' (Watson, 1924). However, with an understanding as the one above, we have a framework that can bring along new discoveries. It says, basically, that even though we should not abandon the metaphysical mind-body problem, a scientific result does not have to solve it in order to be relevant to consciousness studies. Scientific results have in the past been considered irrelevant to consciousness studies because empirical data do not directly answer metaphysical questions. For example, no brain-scanning experiment revealing neural correlates of consciousness would tell us whether those physical correlates are identical to consciousness or not. With a subdivision of the mind-body problem as the one above, one can have progress in consciousness studies without progress in ontological issues.

The methodological interpretation is currently in a state of a 'methodological parallelism'. Even though it may appear dualistic, it is certainly not committed to ontological dualism. It does not claim that consciousness and the brain are in reality separate. As such, I consider it a strength that empirical studies can be carried out without strong ontological assumptions about the mind-brain relation and any changing of this (in case one wishes to get rid of every kind of dualism, hence also a methodological one) will make the value of empirical data dependent on theory. One problem with this would be that, since the ontological debate has not found a final solution over the centuries, and probably will not find one in the immediate future, we would not succeed in having one or just a few empirical accounts of consciousness with generally accepted results. We would have as many empirical accounts as we have theoretical ones, and a solution to the methodological problem would have the same hopeless status as the ontological one, as it would be subordinated to it. Methodological problems related to consciousness studies most certainly exist, especially in relating experimental data to subjective reports, but there are better reasons to be an optimist when it comes to a solution in the near future (see, e.g. Baars, Banks, & Newman, 2003; Depraz, Varela, & Vermersch, 2000; Jack & Roepstorff, 2002; Overgaard, 2003). For example, some recent work has stressed that conscious experiences, although observable from a first-person perspective, are not confined to that person. Experience can be communicated and understood by others who also have their own first-person perspectives. It has thus been argued that consciousness is not some elusive solipsistic island beyond the reach of science but, on the contrary, within reach of any scientist willing to listen to his or her subjects (Jack & Roepstorff, 2003, 2004).

A further gain from this kind of subdivision of the problems of consciousness is that at least it may help avoid a return to the traditional division between philosophical problems and scientific problems. While I hold it to be the case that empirical data have certain and different conditions of validity than does theory or philosophy, it would be very premature and probably wrong to say that they belong to two separate domains. For the methodological interpretation of the mind-body problem, it should especially be clear how philosophical thinking about methodology is connected to the methodology of scientists. Furthermore, certain empirical findings such as 'mirror neurons', discussed below, could be taken as a challenge from the neurosciences to metaphysics, the traditional 'bastion' of philosophy.

### **Readings in consciousness studies**

As I have tried to explain, the problem of consciousness amounts to several distinctly different problems, each of which may be approached in a variety of ways. As we shall now see, the four books under review do differ markedly in what they take to be the problem – or problems – of consciousness. The first book (Zeman, 2003) tries to integrate scientific results into philosophical discussions of the epistemological and ontological problem. The second book (Stamenov & Gallese, 2002) includes hardly any meta-reflections about the problem of consciousness, but it touches upon many methodological discussions that are specifically relevant for theoretical issues of subjectivity, inter-subjectivity and empathy. The third book (Gray, 2004) tries to integrate experimental research

in ontological discussions as does the first book, but with a much stronger emphasis on the empirical results. The book concludes with a defence of a certain neuropsychological theory of consciousness. In contrast to the first three, the fourth book in review (Blackmore, 2003) identifies all three problems of consciousness, yet not always with a clear distinction between them, in the attempt to arrive at a more neutral textbook introducing the field.

In the following, I examine each book more thoroughly in the light of the distinction between the three basic problems: (1) What is the physical status of consciousness? (2) How do we observe consciousness (in ourselves and others)? (3) How should we study consciousness methodologically?

### **Zeman (2003): Consciousness: A user's guide**

Adam Zeman's book, *Consciousness: A user's guide* (2002) is written with an enchantment with the mystery of how all sorts of conscious experiences can be related to physical brain processes – coupled with some pessimism regarding the possibility of ever actually resolving the issues. The book is written in an almost poetic style with several references to how the author himself experiences the world, which sometimes disrupts the clarity of the argument. In other places, however, it makes the presentation more attractive. Zeman seems to endorse an eclectic kind of approach where neuroscience, cognitive science and lines of thinking in the humanities can be integrated in the absence of a larger theoretic framework to defend it.

Zeman starts out by providing a review of the research on the brain mechanisms he finds relevant for debating consciousness in philosophical terms. In particular, he discusses the brain mechanisms involved in being awake and in dreaming sleep, and neural correlates to visual consciousness. He ends up arguing, alongside with what is elsewhere in the air in cognitive neuroscience, that consciousness cannot be located in one or a few areas in the brain (Dennett & Kinsbourne, 1992). Instead, he agrees with Llinas and Ribary (1993) and Crick and Koch (1998), for example, in claiming that visual consciousness appears to be related to rapid synchronized activity between cells that can be physically quite far apart in the brain.

Based on this understanding of the neural basis of consciousness, he suggests that the function of consciousness relates to the possibility of changing one's behaviour when automatic routines fail. Consciousness of a given situation, Zeman argues, helps us to select an appropriate action. Looking at the evidence he includes in his review, as well as his proposed function of consciousness, Zeman may be criticized for conflating consciousness and attention. Whereas the two mental phenomena are often practically related (we are generally conscious of what we are attending to and vice versa), they can be experimentally disentangled. For instance, blindsight patients, who claim to be blind after damage to the primary visual cortex despite preserved visual functions, have been shown to improve in guessing tasks when asked to attend to what is presented to them in the blind part of their visual field. Here, attention can be increased to enhance performance without any change in consciousness.

One attractive feature of Zeman's book is that it does not, unlike several recent texts, aggressively promote any specific theory of mind-body relations. Zeman argues that, even with his proposal concerning a neural substrate and a function for consciousness, we still have not explained consciousness. He has not told us why we actually experience, say, music – why it has a quality of sound. In principle, we could just as well imagine the brain processes involved in listening to music as well as the related cognitive functions to take place in 'phenomenal darkness'. Zeman points out, much like Baars (1988), that the philosophical fight between reductionists and dualists may be a dead end, where the reductionists argue that consciousness is either the same as, or reducible to, brain processes, and the dualists that they are fundamentally different due to some property or aspect of its ontology. Broadly speaking, all known philosophical theories about mind-brain interactions are capable of explaining the existing empirical observations.

Zeman's line of argument is generally sound and his conclusion is not revolutionary. To some degree, it may remind the reader of the pessimistic conclusions of Colin McGinn, for example, that human beings do not have the cognitive capacity to answer the ontological problem (1989). It is

however also important to keep in mind that Zeman's pessimistic view on solving the problem of consciousness relates specifically to the 'ontological problem of consciousness'. As soon as Zeman refers to empirical investigations, his pessimistic conclusions no longer apply.

### **Stamenov and Gallese (2002): Mirror neurons and the evolution of brain and language**

For some researchers, the discovery of so-called 'mirror neurons' is an example of how empirical data may in fact inform ontological discussions about consciousness. Mirror neurons respond to the observation of actions of others in both humans and non-human animals. They 'mirror' an action in such a way that areas of, say, motor cortex are activated 'as if' the subject was in fact himself performing the activity he was observing. Therefore, activation of a mirror neuron system has been suggested as the neural correlate of empathy and the understanding of others (Gallese, 2001). For technical and ethical reasons, mirror neurons have been studied most extensively in monkeys, but have also been observed in humans.

The acquisition of new functions supported by corresponding brain circuits influences the performance of the mirror neuron system not only at a macro-scale, but even at single neuron performance. Thus, we seem to be speaking of a phenomenon independent of higher-order functioning. Now, according to some theoretical accounts, consciousness is a specifically human, higher-order mental function (Rosenthal, 1993) that somehow monitors lower-order mental functions, such as sensations. Some theorists go further and claim that it is specifically these higher-order functions that allow us to understand the actions of other people and communicate with them about it (Fodor, 1983). This is sometimes referred to as 'theory theory', or the idea that we understand others on the basis of a (rather advanced) theory about their thoughts and feelings. Such an approach to the nature of consciousness is brought into question by the discovery of neurons tuned to fire not only to the enaction but also the observation of certain kinds of behaviour (Rizzolatti, Fadiga, Gallese, & Fogassi, 1996). The argument goes that if mirror neurons may be interpreted as a neural correlate of empathy, in a very basic form, and if mirror neuron systems are not dependent upon consciousness, we cannot argue (1) that consciousness is a higher-order phenomenon nor (2) that understanding of others is dependent on consciousness.

Thus, the discussion of mirror neurons links to core discussions in the ontological debate about consciousness, suggesting that we must either turn away from higher-order representation theories in the tradition of Rosenthal, Carruthers or Lycan, or we must see empathy as a completely unconscious phenomenon. Since the latter is intuitively hard to embrace, mirror neurons easily fit into existing theoretical attempts to go against higher-order theory.

As previously mentioned, it has been argued that first-person perspectives are not solipsistically confined to individuals: Experience can be communicated and understood by others who also have their own first-person perspectives (Jack & Roepstorff, 2003, 2004). This approach fits into the mirror neurons debate, as it suggests we have the understanding of other people as an innate capacity that under normal circumstances 'comes along' with having a first-person perspective. Mirror neurons, intuitively at least, look more like a neural substrate to a 'direct understanding' of other people rather than a neural substrate to some separate cognitive module subserving the same functional role.

The first part of *Mirror neurons and the evolution of brain and language* discusses the functional role of mirror neurons and attempts to define cortical circuits for action understanding and imitation, drawing upon evidence from behavioural studies as well as brain-imaging techniques and transcranial magnetic stimulation. The second part goes on to discuss the mirror neuron system in much more detail. Gerhard Roth contributes with a thought-provoking paper, showing that all popular beliefs about the uniqueness of the human brain compared with the brains of other animals are either misconceived or exaggerated and based on much too little evidence. Later in the same part, Ava Senkfor demonstrates that different brain circuits are engaged depending on prior experience when performing an action when watching an action being performed and when imagining an action. Her results indicate a high degree of specificity of

memory traces, and they display some quite interesting aspects of mirror neuron functioning, for example that the same neurons are active when the participants move their own right hand and when they observe the experimenter moving his right hand (despite the fact that they faced each other). McGlone, Howard and Roberts use fMRI to look for the location of mirror neurons in humans more precisely. Their results indicate that there is a localizable mirror system network, involving both motor and sensory areas.

The third part of the book focuses on the evolution of the brain, language and communication. Not least, this part attempts to show the relevance of mirror neurons in the discussion of empathy and consciousness. Edda Weigand points out that mirror neurons integrate different dimensions of cognition. For instance, she claims that language was made possible in evolution by way of the activity of mirror neurons. Stamenov argues, somewhat in opposition to Roth (above), that, whereas the mirror neuron system supposedly is the same in monkeys and in humans, it cannot account for human higher cognitive functions that differ from the cognitive functions of monkeys. Bernard Bichakjian proposes that evolution of speech and evolution of language have followed two distinct routes through phylogenesis. Speech articulation, Bichakjian argues, is controlled by Broca's area whereas the areas associated with verbal intelligence fall largely outside of Broca.

The fourth part of the book considers some applications of the findings. Billard and Arbib suggest the use of computational neuroscience to simulate the neural activations of mirror neurons underlying imitation to better understand levels of imitation in animals and humans. Womble and Wermter develop a model of syntax acquisition, showing that a connectionist network has difficulties in fully learning a grammar without the use of an abstracted system of simulated mirror neurons.

The book presents a diversity of discussions and an overview of research on mirror neurons. As with other books on recent and fascinating discoveries, there is at times a tendency to wild speculation about how many diverse phenomena may be explained by mirror neurons. Indeed, in their introduction, the editors claim that 'it is hard to overestimate the importance of this discovery'. While this approach of integrating neuroscientific discoveries into philosophical debates on consciousness could be potentially very enriching, the book hardly explains exactly why the editors and contributors believe that the study of mirror neurons will promote the progress of consciousness studies. The prospects of the book in this regard seem not yet fully explored, although some philosophical frameworks are likely to pick up on the discussion just as it happened in the discussion of higher-order theory in regard to the mirror neuron phenomenon as such. One thing especially missing in the book is an explicit discussion of the correctness of the assumption that mirror neurons must be the neural correlate of empathy. If this were somehow falsified, the relevance of mirror neurons to consciousness studies would clearly be in question.

### **Jeffrey Gray: Consciousness**

The late Jeffrey Gray was an influential figure in British neuropsychology, and his last book is about the ontological problem of consciousness. Gray had already written about consciousness in 1971, long before it became a fashionable topic, and he then criticized philosophical theories for not formulating the problem of consciousness as a matter of relating consciousness to neuroscience and to psychology. Now, however, Gray finds almost the opposite to be the case and says that the progress of many recent experiments addressing consciousness has put us in danger of overlooking deep conceptual issues which also need to be addressed if we are to fit consciousness into the framework of natural science. In light of the discussion above, I would add to Gray's critique raised in 1971 that the ontological question about consciousness can be set aside and dealt with separately. Thus, although I very much agree with Gray that the further question about the status of consciousness in neuroscience is important to resolve, it is a methodological question and thus not necessary for addressing the ontological question.

Gray succeeds in providing a good overview of many of the empirical findings in the consciousness literature, although the choice of experiments for review is, of course, selected based on his line of argumentation. Nevertheless, the overview gives very good insight into

psychological and neuroscientific consciousness research, and for this reason, in particular, I can highly recommend this book. A further reason for recommending this book is the exceptional expertise Gray shows in both scientific and philosophical thinking, which is rather unusual even in an interdisciplinary field.

In the course of reviewing the empirical findings, Gray identifies a paradox. The existing evidence seems to suggest both that consciousness has no immediate causal powers (e.g. Libet, 1985) and that it is not an epiphenomenon, i.e. not devoid of causal powers. Instead of adopting and defending either of these two opposing positions, Gray attempts to reconcile them. He suggests that consciousness serves as a 'late detector of error'. Consciousness, he claims, has the function of monitoring brain activity, comparing results with given goals and resetting the system to correct errors and avoid future errors. This hypothesis, Gray claims, satisfies both claims in the empirical literature because it grants causality to consciousness, but only 'long-term causality'. Thus, Gray's take on the ontological issue leaves us with an unusual version of functionalism, where consciousness is identical to a higher-order, functional property of the brain.

One cannot, however, help feeling a bit confused about Gray's account as a theory about consciousness. It is certainly not impossible that the brain might possess a 'comparator system' involved in error detection. What, however, remains a mystery is why it is just this system that should be related to conscious experience. After all, from a phenomenological point of view, we experience not just action decisions, but also sensations, emotions and thoughts. From this perspective, we find no evidence for a 'special role' for consciousness in neural structures specifically related to decision making and error monitoring. This is a standard flaw in many scientific books on consciousness (Overgaard, 2004). The authors assume a relation between consciousness and a specific cognitive or neurological function, then give an in-depth description of these functions, and finally conclude that they have explained consciousness. This line of argument is, however, no stronger than the proposed link between consciousness and the relevant function.

In this particular book, the confusion has a further origin. Gray seems to raise a methodological question: how does consciousness fit into the neurosciences? However, he also makes several claims about 'what consciousness is', and to an extent that makes the ontological question seem the primary concern of the book. His ontological commitment to a comparator system will, to some theorists, sound like a philosophical alliance with the abovementioned higher-order theory of David Rosenthal (among others), which says that consciousness happens when one unconscious thought has another mental state (say, a perception) as its intentional content (Rosenthal, 1990). Thus, some might think this book provides empirical evidence for this line of philosophical theory. It is, however, fundamental to keep in mind that the comparator system hypothesis is more theoretical than directly observed in experimental results, and the relation between this proposed system and consciousness is even more a theoretical speculation.

### **Susan Blackmore: Consciousness: An introduction**

Susan Blackmore's *Consciousness: An introduction* is the first attempt to write a scholarly textbook on consciousness studies. It would seem to be aimed at undergraduate students, although other academics unfamiliar with the 'consciousness literature' may benefit from reading it as well.

The book is organized into nine sections, each comprising three chapters. The sections are entitled 'The problem', 'The world', 'The self', 'Evolution', 'Artificial consciousness', 'The brain', 'Borderlands', 'Altered states of consciousness' and 'First person approaches'. Each chapter is designed to be read on its own, allowing the reader to look at arguments and scientific evidence concerning particular aspects of consciousness studies. As such, the book lives up to the requirement of a good textbook in being both popular and scientifically sound at the same time.

Anybody seriously involved in 'consciousness studies' will, no doubt, feel irritated that certain aspects of the debate are ignored or not explored thoroughly in depth. However, I, at least, cannot avoid admiring how Blackmore succeeds in balancing the text so it never becomes too technical

for the inexperienced reader nor so shallow or superficial that the 'professionally involved' would have to object to it.

A special aspect of the book is the inclusion of little 'consciousness exercises' to illustrate the text. Again, this succeeds in illustrating, among other things, that the contents of our consciousness at any given moment are not as easily accessible as we might intuitively think. The little exercises also make the read more enjoyable.

The book stands out from all others in the field in that it attempts to take on an objective and neutral attitude to the theories and experiment it reviews. However, in her discussion of philosophical dualism, Blackmore suddenly changes her style and verges on hostility. While this attitude may be rather standard among modern-day psychologists, it spoils an otherwise successful attempt to provide a non-partisan book about consciousness. Nevertheless, the book is well worth using for an undergraduate course on consciousness. In addition, the book stands apart from the others in touching upon all interpretations of the problem of consciousness – ontological and epistemological as well methodological. This is, of course, also necessary for a book trying to encompass the entire field. The book, however, makes no clear attempt to spell out the differences between the different kinds of problems. One could criticize the book for not being sufficiently explicit in pointing out when arguments have an ontological, epistemological or methodological aim.

### Discussion and comparison

Many books on consciousness have already appeared. Many have been enriching, educating and thought-provoking in discussing philosophical issues in relating mind to brain or in discussing scientific evidence for, say, the dissociation of conscious and unconscious processes or neural correlates of consciousness. Good introductory texts that outline different aspects of problems related to consciousness studies are Chalmers (1996) or Metzinger (1995) for strictly philosophical approaches. For more scientific texts, Baars *et al.* (2003) is especially worth consulting. For books attempting a theoretical approach based on neuroscience and cognitive science, one may consult Cotterill (1998) and Marcel and Bisiach (1988) or, if one is interested in the practice of phenomenology in cognitive science, I would recommend Petitot *et al.* (1999).

The books I have been reviewing are a welcome addition to this literature. Zeman contributes to the long-standing controversy of the relation of mind to body by appeal to neurological sources of evidence. While his line of argument is a serious contribution to the debate about mind-body theory, the book does not represent a uniquely new theoretical position. Gray also reviews recent neurological and psychological evidence and discusses it in the light of philosophical theories about consciousness. Nevertheless, he ends up in opposition to Zeman by concluding that not only is it possible to solve the ontological problem of consciousness, it is already at this stage of scientific progress possible to explicate and defend one specific theory. Thus, Gray defends an ontological theory about consciousness being a 'monitor and veto system' in relation to action control.

The controversy between Zeman and Gray serves as a nice illustration of typical difficulties in consciousness studies. Even when confronted with identical evidence, thinkers seem to come to very different conclusions. Seemingly, there is a relatively long distance from data to theory when it comes to measures of brain and neural activity, which is a signal that the field as a scientific enterprise is still very undeveloped. The same problem is manifest in Staminov and Gallesse although the symptom is different. Here, as discussed in the review, the theoretical link to consciousness is far from fully explored, since the book takes a more empirical perspective than the other two.

While many empirical scientists identify the problem of consciousness with the ontological interpretation, there are, as I have been arguing, actually several different problems of consciousness. Although they are not interrelated by necessity, one will need methodological and conceptual insight to say how definitions of consciousness relate to results from neuroscience.

Blackmore's *Consciousness* stands out from the others in her attempt to encompass consciousness studies in a single text from a theoretically neutral perspective. Although her book

does not present any new theory or data (since it is a textbook reviewing the field), it could be seen as the most uniquely new contribution to the field of the four books here reviewed, since it raises the question of where consciousness studies should go next.

### Consciousness studies and psychology

In Blackmore's book, we find an implicit answer, at least, to the question raised earlier in this review: should consciousness be the subject of a discipline on its own with university courses and a specific science devoted to consciousness? Or should it rather be a perspective integrated in disciplines such as psychology and neuroscience? Blackmore's book, not just in its text but also in its format, seems to favour the first solution.

I shall not here go into many details regarding the paradigmatic advantages or disadvantages of this suggestion to a solution of the 'disharmony' between cognitive science methods and theory, and consciousness. I shall, however, give a personal view on the matter. Arguments in favour of treating consciousness studies as separate from other sciences would be (1) the failure, generally speaking, of cognitive science to study consciousness, and thus properly to distinguish between conscious and unconscious processes and (2) that consciousness studies constitutes an interdisciplinary field that does not fit into current frameworks to be found, for example, in psychology.

However, one could also argue that difficulties within one discipline in being all-encompassing of its subject matter form a rather poor argument for establishing a whole new science. Although cognitive scientists have been uncomfortable discussing consciousness, they would presumably not be happy being known as 'scientists of the unconscious' either. Furthermore, it is not as though other concepts in psychology are less potentially interdisciplinary than is consciousness. Intentionality, selfhood, personality, volition and many other standard concepts in psychology all carry much philosophical baggage. Also, other fields of psychological research and theory lean on neuroscience, philosophy, anthropology, sociology and many other disciplines. It is my personal belief that the most benefit would come out of the current broad interest in consciousness if the insights from consciousness studies could be integrated into psychology, neuroscience and other disciplines.

This integration is not just to be considered as 'the way forward' for consciousness studies, but for psychology as well. Since the fall of introspectionism in the early part of the twentieth century, the topic of consciousness has been met with great scepticism by the majority of especially experimental psychologists. Especially, the methodological debate about consciousness holds the promise of reconciling some of the abandoned concepts such as subjectivity, intentionality and experience with leading branches of experimental psychology, i.e. cognitive science and cognitive neuroscience. Other parts of the ongoing debate on consciousness can also inspire psychologists to take consciousness seriously, even though a 'solution' may not be just around the corner. If psychologists will not accept their fate as 'scientists of the unconscious', unable and unwilling to discuss the one aspect of our minds that is the most present to all human beings – that they have conscious experiences – this is the right time to consider how consciousness can be brought back on to the psychological agenda.

It seems likely that many more publications along the lines of the reviewed books will appear in the future. While such are still needed, I hope this review essay will encourage future authors to show us how to integrate into the existing sciences the still growing body of knowledge and ideas in consciousness studies. Whether this will result in dramatic paradigmatic changes or in no changes at all for the individual science, it would, so to speak, be the way out *and* forward for consciousness studies.

**MORTEN OVERGAARD** (Hammel Neurocenter, Aarhus University Hospital, Denmark)

## References

- Allport, A. (1988). What concept of consciousness? In A. Marcel & E. Bisiach (Eds.), *Consciousness in contemporary science* (pp. 159–182). Oxford: Oxford University Press.
- Baars, B. J. (1988). *A cognitive theory of consciousness*. Cambridge: Cambridge University Press.
- Baars, B. J. (2003). The mind-body problem is scientifically untestable and irrelevant. *Bulletin fra Forum for Antropologisk Psykologi*, 13, 32–34.
- Baars, B. J., Banks, W. P., & Newman, J. B. (2003). *Essential sources in the scientific study of consciousness*. Cambridge, MA: MIT Press.
- Block, N. (1978). Troubles with functionalism. In C. W. Savage (Ed.), *Perception and cognition* (pp. 261–325). Minneapolis, MN: University of Minnesota Press.
- Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences*, 18, 227–287.
- Chalmers, D. J. (1996). *The conscious mind*. Oxford: Oxford University Press.
- Chomsky, N. (1957). *Syntactic structures*. The Hague, Netherlands: Mouton.
- Chrisley, R. (2001). A view from anywhere: Prospects for an objective understanding of consciousness. In P. Pykkänen & T. Vaden (Eds.), *Dimensions of conscious experience* (pp. 3–13). Amsterdam, Netherlands: John Benjamins.
- Chrisley, R. (2002, May). *Is there a seemings-is distinction for seemings?* Paper presented at the First Meeting of the Nordic Network for Consciousness Studies, Skövde, Sweden.
- Churchland, P. S. (1986). *Neurophilosophy*. Cambridge, MA: MIT Press.
- Cotterill, R. M. J. (1998). *Enchanted looms*. Cambridge: Cambridge University Press.
- Crick, F., & Koch, C. (1990). Toward a neurobiological theory of consciousness. *Seminars in the Neurosciences*, 2, 263–275.
- Crick, F., & Koch, C. (1998). Consciousness and neuroscience. *Cerebral Cortex*, 8, 97–107.
- Dennett, D. C., & Kinsbourne, M. (1992). Time and the observer: The where and when of consciousness in the brain. *Behavioral and Brain Sciences*, 15, 183–200.
- Depraz, N., Varela, F. J., & Vermersch, P. (2000). The gesture of awareness: An account of its structural dynamics. In M. Velmans (Ed.), *Investigating phenomenal consciousness* (pp. 121–136). Amsterdam: John Benjamins.
- Fodor, J. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.
- Gallese, V. (2001). The 'shared manifold' hypothesis: From mirror-neurons to empathy. *Journal of Consciousness Studies*, 8, 33–50.
- Jack, A., & Roepstorff, A. (2002). Retrospection and cognitive brain mapping: From stimulus-response to script-report. *Trends in Cognitive Science*, 6, 333–339.
- Jack, A., & Roepstorff, A. (2003). Why trust the subject? *Journal of Consciousness Studies*, 10, vi–xx.
- Jack, A., & Roepstorff, A. (2004). Trust or interaction? In A. Jack & A. Roepstorff (Eds.), *Trusting the subject* (Vol. 2). Exeter, UK: Imprint Academic.
- Jackson, F. (1986). What Mary didn't know. *Journal of Philosophy*, 83, 291–295.
- James, W. (1890). *Principles of psychology*. New York: Holt.
- Kripke, S. (1980). *Naming and necessity*. Cambridge, MA: Harvard University Press.
- Leahey, T. (1987). *A history of psychology: Main currents in psychological thought*. Englewood Cliffs, NJ: Prentice-Hall.
- Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and Brain Sciences*, 8(4), 529–566.
- Llinas, R., & Ribary, U. (1993). Coherent 40-Hz oscillation characterizes dream state in humans. *Proceedings of the National Academy of Sciences*, 90, 2078–2081.
- Marcel, A. J., & Bisiach, E. (1988). *Consciousness in contemporary science*. Oxford: Clarendon Press.
- McGinn, C. (1989). Can we solve the mind-body problem? *Mind*, 98, 349–366.

## Copyright © The British Psychological Society

Reproduction in any form (including the internet) is prohibited without prior permission from the Society

438 Book review essay

- McGinn, C. (1991). *The problem of consciousness: Essays toward a resolution*. Oxford: Blackwell.
- Metzinger, T. (1995). *Conscious experience*. Exeter, UK: Imprint Academic.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*, 81-97.
- Miller, G. A. (1962). *Psychology: The science of mental life*. New York: Harper and Row.
- Miller, G. A. (1979). *A very personal history*. Cambridge, MA: MIT Press.
- Miller, G. A. (2003). The cognitive revolution: A historical perspective. *TRENDS in Cognitive Sciences*, *7*, 141-144.
- Nagel, T. (1974). What is it like to be a bat? *Philosophical Review*, *83*, 435-451.
- Nagel, T. (1986). *The view from nowhere*. Oxford: Oxford University Press.
- Nelkin, N. (1993). What is consciousness? *Philosophy of Science*, *60*, 419-434.
- Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.
- Overgaard, M. (2003). On the theoretical and methodological foundations for a science of consciousness. *Bulletin fra Forum for Antropologisk Psykologi*, *13*, 6-31.
- Overgaard, M. (2004). Confounding factors in contrastive analysis. *Synthese*, *141*(2), 217-231.
- Petitot, J., Varela, F., Pachoud, B., & Roy, J. M. (Eds.). (1999). *Naturalizing phenomenology*. Stanford, CA: Stanford University Press.
- Putnam, H. (1967). The nature of mental states. In W. H. Capitan & D. D. Merrill (Eds.), *Art, mind, and religion*. Pittsburgh, PA: Pittsburgh University Press.
- Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, *3*, 131-141.
- Rosenthal, D. (1990). *A theory of consciousness* (Report No. 40). Bielefeld, Germany: Universität Belefeld, Zentrum für Interdisziplinäre Forschung.
- Rosenthal, D. (1993). Higher order thoughts and the appendage theory of consciousness. *Philosophical Psychology*, *6*, 155-165.
- Searle, J. (1980). Minds, brains and programs. *Behavioural and Brain Sciences*, *3*, 417-457.
- Searle, J. (1992). *The rediscovery of the mind*. Cambridge, MA: MIT Press.
- Shallice, T. (1972). Dual functions of consciousness. *Psychological Review*, *79*, 383-393.
- Sperling, G. (1960). Afterimage without prior image. *Science*, *131*, 1613-1614.
- Watson, J. B. (1924). *Psychology: From the standpoint of a behaviourist*. Philadelphia, PA: J.B. Lippincott.
- White, P. (1980). Limitations on verbal reports of internal events: A refutation of Nisbett and Wilson and of Bem. *Psychological Review*, *87*, 105-112.

Copyright of British Journal of Psychology is the property of British Psychological Society and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.