

## *Book Reviews*

**William Hirstein**

*Brain Fiction*

Cambridge MA: MIT Press, 2005, 289 pp. (hbk.). ISBN 0-262-08338-8

*Reviewed by John Bickle*

*University of Cincinnati*

Confabulation — the seemingly sincere utterance of (sometimes) plausible, contextually relevant falsehoods — has long intrigued philosophers. *Something* seems philosophically interesting about a blind patient who claims to see and readily answers questions (incorrectly) about his nearby visual space, or an amnesic who tells an elaborate, detailed story about last night's escapades with companions, some perhaps, long dead, even though he was safely tucked into bed by early evening. Yet these 'gee whiz' phenomena have not yielded much detailed philosophical impact.

William Hirstein seeks to change this. He argues that a detailed understanding of the clinical symptoms, a survey of the brain damage involved and a philosophical analysis built upon these neurological details reveals that confabulation is an epistemic disorder — 'a malfunctioning of epistemic processes'. Understanding it thus tells us important things about mind, knowledge, and human nature. He speaks with authority, given his combined backgrounds in neurophilosophy and clinical neuropsychology.

In accord with best practice in neurophilosophy, Hirstein holds off on a 'definition' of confabulation (Chapter 8) until he has surveyed the relevant scientific data. After a stage-setting introductory chapter and a brief account of the neurophilosophy he will be using, Hirstein leads the reader through confabulation conjoined with memory deficits (Chapter 3), sociopathy (Chapter 4), misidentification disorders (like Capgras' syndrome, where patients come to believe that a loved one has been replaced by a look-alike imposter — Chapter 5), anosognosia (denial of illness — Chapter 6), and commisurotomy ('split-brain' patients — Chapter 7). Each chapter contains an accessible overview of the basic clinical symptoms and the known sites of associated brain damage. Some anatomical commonalities emerge. Confabulators have damage to (orbital) frontal lobes or

densely connected areas, in addition to damage in areas subserving their primary cognitive deficit (memory, sensation, motor). Numerous interesting points emerge in Hirstein's discussions, including the role of 'mind reading' problems (an inability of confabulators to notice or acknowledge disbelief in their listeners), and the observation that disorders like obsessive-compulsive disorder and hypochondria are epistemic 'mirror images' to some varieties of confabulation.

Chapter 8 tries to deliver the philosophical goods. Confabulatory phenomena, both clinical and everyday examples, are unified under the rubric of 'self-knowledge deficits.' Brains contain both 'representation-creating' and 'representation-checking' processes. Confabulation begins with the first process. The creative process produces claims with 'unacceptably low truth probabilities', then fails again: 'next, something causes the brain to miss its second chance to get it right — it fails to detect the defective claim'. Confabulators lack or have a damaged 'checking process', a representation-evaluating process coupled with a behaviour-inhibiting process. According to Hirstein, his account handles the symptoms of clinical and everyday confabulation, is consistent with the known brain damage, and squares with relevant phenomenology (subjective judgements of the falsity or implausibility of our thoughts). It also provides for confabulation's epistemic features (pages 187–98). Confabulatory claims are 'ill-grounded', the confabulator 'does not know' but 'should' have known (had her mind been functioning normally) that her claims are ill-grounded, yet is 'confident' that they are accurate. 'In confabulation, the mind is working suboptimally; processes that evolved to have quite definite functions are not operating.' Hirstein tries to connect his discussion with ongoing 'naturalization' projects in current epistemology, with reliabilism, and with the epistemic internalism/externalism debate.

Chapter 9 applies this definition to the vexing philosophical issue of understanding self-deception non-paradoxically. Chapter 10 briefly speculates on the role of confabulation in normal human life, ending with questions about implications for other topics in philosophy, including consciousness, the (narrative) self and rationality.

Readers of this journal might be especially interested in Hirstein's take on the issue of confabulated reasons for voluntary actions (pages 171–5). 'Skeptics' like Michael Gazzaniga and Daniel Wegner see dire consequences in this research for our notions of a free conscious will and our everyday explanations of human action. 'Believers', on the other hand, insist that sometimes (not always!) conscious aspects of brain function do guide behaviour. Hirstein is a Believer. 'What the skeptics are describing are pathological states, or at least states in which the human brain is being used in suboptimal ways.' He puts his two-stage epistemic account of confabulation to work: 'Consciousness functions as a sort of testing ground, where flawed representations are prevented from taking part in the causation and guidance of action.' He applies his intriguing idea that obsessive compulsive disorder is the epistemic and clinical mirror syndrome of confabulation: 'If conscious states have no real effects, why are the obsessive thoughts so incredibly distracting and debilitating?'

Hirstein has packed a wealth of intriguing scientific information into this book. Even if the philosophy can seem a bit thin in places, readers will be intrigued by his detailed surveys of clinical confabulatory symptoms and will learn something about the state of current neuropsychology and neurology. Yet, in the final analysis, those with some knowledge of basic neuroscience may also feel short-changed. In Chapter 8, Hirstein blithely remarks that ‘the brain is filled with representations’, some ‘imagelike’, others ‘more conceptlike’. Hmm. To my (admittedly reductionistic) view, the brain is filled with neurons and glial cells: with collections of membrane-bound cellular and molecular tricks for getting ions selectively across those membranes. Any account of cognition that claims ‘neurobiological plausibility’ has to connect with that level of description, but Hirstein hasn’t done so (yet). Claims like ‘the brain has a large network of interconnected structures, and sites of damage relevant to confabulation are part of this network[;] the orbitofrontal cortex, the areas around the superior temporal sulcus, and the inferior parietal cortex are some of the areas affected’ (p. 181) *don’t propose mechanisms*. Hirstein’s own wording reveals this approach’s limitations. Here is his account of judging something to be false or improbable: ‘I consider the thought . . . I hold it in consciousness, and then, *rather miraculously*, if any other representation in my brain conflicts with it, it speaks up’ (p. 182, my emphasis). But given what we already know about the physiology of neurons in these regions, it is reasonable to speculate that Hirstein’s epistemic concepts will be left behind as further neurophysiological discoveries are made. Where such discoveries will leave the philosophy of clinical neurology remains an open question.

### **Susan Blackmore**

*Consciousness: An Introduction*

Hodder & Stoughton, 2003, 460pp, £19.99, ISBN 0340809094

*Reviewed by Tim Calton*

*University of Nottingham*

It’s not easy to present the many complexities of contemporary consciousness studies to an interested lay audience (of whom I am one). The task is complicated by lingering dualism that pervades our culture (a particularly recalcitrant meme perhaps?). It is to Susan Blackmore’s credit that she is open about the potential for perplexity in such a collision of epistemologies. She deals with it well in her first few chapters. At 460 pages the book is no lightweight, but in fact proved an easy read due to an harmonious conjunction of demotic text and images, pop biographies of leading luminaries in the field and active encouragement to engage in ‘practice’ and ‘activity’ tasks (of which more later). No doubt this is intended to both allay fear of the unknown and pique interest, though I can imagine that it might provoke raised eyebrows in some academic quarters. But then the book is not intended for professional students of consciousness; it’s one for the neophyte.

There are nine main sections of three chapters each. The first six sections (encompassing discussion of the main theoretical problems, the elision of the world and the self, the evolution of human consciousness and the potential for artificial consciousness, and reductionist theories) are essential. The final three spiral gently off into more contentious (and by dint of that at times more interesting) territory, ending on a rather declamatory exposition of the benefits of Buddhism. Apart from an intriguing but far too brief discussion of phenomenology, these last few sections feel rather superfluous.

The first section introduces most of the historical problems, familiar to any philosopher of consciousness but nevertheless intriguing to newcomers. The central position occupied by Chalmers's 'hard problem' is delimited and reviewed, before the important question of what consciousness might be for is posited. At this point I began to understand the rhythm of the book; the push and pull of question and response, much in the Socratic tradition. However each successive chapter left me feeling somewhat unfulfilled, perhaps because the present state of the art in consciousness studies inevitably abandons us in a morass of seemingly unanswerable questions.

The second section holds the attention with accounts of Libet's counterintuitive empirical work on the timing of subjective experience and of various eminent thinkers putting the intellectual boot into dualist theorising. I caught echoes of the destruction wrought by Ryle in Dennett's exhortation to demolish the 'Cartesian theatre', and it was no surprise to learn that he was one of Ryle's PhD students. This intellectual iconoclasm continues in the third section which addresses the problem of identity and the self. I suspect most consciousness studies neophytes will be familiar with the notion of multiple personalities, but fewer may sit comfortably through the discussion of the ramifications of split-brain experiments or the entire chapter devoted to theories of self — reading that one (?) should regard oneself as a 'centre of narrative gravity' could be a rather intimidating introduction to the vagaries of late-modern thought. Equally the discussion of agency and free will, again drawing heavily on Libet's classic studies, stridently challenges 'common sense' conclusions about free will. Here Blackmore once again employs Dennett's philosophy contrapuntally and, I felt, with undue veneration for it.

The fourth section opens with a relatively straightforward exposition of evolutionary theories, before an exercise in speculation entitled 'the function of consciousness' (including a brief recap of Blackmore's own work on memes), concluding with a chapter on animal minds. It then segues into a section devoted to the debate about artificial consciousness which, via Turing's Test and Searle's Chinese Room, arrives at Penrose and Hameroff's other-worldly treatise on quantum consciousness — here Blackmore's hitherto confident grasp of her subject matter seems to slip somewhat. Section six focuses on the brain, opening with a teasing look at the neural correlates of consciousness, before encompassing chapters on the unity of consciousness and the, often paradoxical, effects of brain damage. My thinking became recursive at this point and I wondered if restating the 'hard problem' might have provided a useful touchstone in this context.

And thus to the last three sections: No doubt the paranormal, the effects of psychotropic drugs and Buddhism are interesting topics, but are they representative of what matters most to students of consciousness? If this book is intended as a taster for the whole field, this question surely has to be asked. For me the answer is no, but I accept that their inclusion may help to induce some (young) people to actually read this book in the first place. Utilitarianism thus won me over, though the imposition of the author's own particular proclivities did grate on me.

In conclusion, then, this book fulfils its purpose admirably in that it describes the current state of the debate about consciousness in both breadth and (some) depth. It does this in a manner conducive to drawing in the newcomer (in particular those at undergraduate level), with ample use of practice, activity and self-assessment exercises. I would comment that some of the practice and activity exercises seemed rather non user-friendly, though perhaps a little ontological discomfiture might be stimulatory rather than inhibitory in this context. From my, albeit rather limited, perspective the only major bone of contention would be the inclusion of the last three sections. Thinking comparatively, I found this book less well written than, say, George Graham's *Philosophy of Mind: An introduction* (which covers much of the same territory), but more up to date and diverse in its purview. All in all this is a worthy effort, and it has clearly set the standard for similar future texts on this most contentious yet exciting of subjects.

### **Eric Dietrich and Valerie Gray Hardcastle**

*Sisyphus's Boulder : Consciousness and the Limits of the Knowable*

John Benjamins, 2004, 133pp. ISBN 90 272 5196 7

*Reviewed by Hugh Noble*

This book is volume 60 in the Advances in Consciousness Research series. It is quite short and written in a clear brisk style which makes for easy reading. I found it both enjoyable and thought provoking.

The approach taken is unusual. Two schools of thought are identified. These don't represent the traditional schism between people who think that consciousness is a physical attribute of the brain and others. Instead we have what the authors dub the '*Naturalists*' and the '*Mysterians*'. These classifications bring together surprisingly diverse groups of bedfellows. The Naturalist group embraces materialists, eliminativists, idealists and even some dualists. They are united only in that they believe that consciousness will be one day explained by a robust scientific theory. Mysterians hold that science is powerless when confronted by consciousness. According to Dietrich and Hardcastle, both groups are wrong.

They avoid the apparent contradiction by making a distinction between acquiring an 'explanation' and 'having a science'. An explanation, they say, entails detailed, intuitively satisfying description at the level of cause and effect. To have a science of consciousness, on the other hand, requires only that we delineate a set of empirical laws governing the correlation between mechanistic

functions of the brain and associated (reported) phenomenal experiences. They claim ‘explanation’ to be impossible, while ‘having a science’ is entirely feasible. The authors acknowledge the existence of a third group — those who think they already have an explanation of consciousness. Dennett’s *Consciousness Explained* is quoted as an example, but his views are dismissed with a one-line expression of incredulity — ‘Such researchers have not been paying attention.’ That is not what I would describe as a well argued case.

The book is in three parts. The first explains what is wrong with the Naturalist view; the second what is wrong with the Mysterians’ approach. The third part discusses implications of the authors’ theory for the future of philosophy. There is also an appendix dealing with David Chalmers’ ideas about the conceivability of zombies. In it they claim to refute his theory relating to zombie-twins — as distinct from general zombie-ism, (on which they remain neutral).

The argument against Mysterians in part 2 is concerned with the nature of scientific explanations, while in part 3 the main conclusion is that many important philosophical questions are undecidable. Since I am someone whom the authors would classify as a Naturalist of the reductionist variety — and one who has, in addition, not been paying attention — I will focus my comments on part 1. Here they claim that there can never be an intuitively satisfying scientific explanation of consciousness because the Cartesian and the zombie intuitions ‘de-rail’ any such attempt. The authors claim to be agnostic on the validity of these intuitions, but say that these are so strong, so widespread and so unshakeable, that they will always make a scientific explanation of consciousness seem unsatisfactory.

Early in the text, the concept of supervenience is introduced and explained. Two examples are given: (1) The temperature of a substance supervenes on molecular agitation; (2) The forward movement of a car supervenes on the mechanical action of its engine and transmission system.

I think this is a mistake. The term ‘supervenience’ has been used by different philosophers in several different ways (see, for example, Johansson, 2002). What is quite clear, however, is that the two examples above are in different categories. The temperature of a substance and its molecular agitation are actually the same thing (viewed from different perspectives). You cannot have one without the other. There is, therefore, no way that the connection between them can be broken. In contrast, the connection between a car’s movement and the operation of its engine is easily broken by placing the car’s wheels in soft mud, for example. The authors mention this difference, but seem unaware of its importance for their own argument. Zombie-ism is central to this argument. However, if consciousness supervenes on brain mechanisms in the same way that temperature supervenes on molecular agitation, then the separation of consciousness from brain operation is impossible and zombie-ism is also impossible. If consciousness is causally connected to mental operations (as in the car example) then the possibility of zombie-ism cannot be excluded, at least in the context of this particular argument.

The rest of the book makes frequent reference to the concept of supervenience, but uses the car example as its paradigm. No justification is given for ignoring

the possibility of the other kind of supervenience. The argument revolves around the need to find explanations (or empirical rules) relating to causal connections. Naturalism is found wanting by these authors because, they say, it cannot provide intuitively satisfying explanations of these causal connections. The materialist approach, by being lumped in with the naturalist group, is subject to the same criticism. However, I suggest that causal connections of the kind they envisage, have no place within a materialist account of things.

### Reference

Johansson, I. (2002), 'On Armstrong's and Lewis' concepts of supervenience', *Nordic Journal of Philosophy*, 3 (1), pp. 119–22.

### Ted Honderich

*On Consciousness*

Edinburgh University Press, 2004. 230 pp (hbk) ISBN 0 7486 1842 2.

*Reviewed by Paavo Pykkänen*

*University of Skövde*

There is a range of convincing arguments to do with the impossibility of explaining consciousness within the framework of orthodox neuroscience. While some criticise these arguments and hope to account for consciousness in terms of conventional science, others explore more exotic alternatives such as panprotopsyism, quantum consciousness, antirealism or dual aspect theory. In *On Consciousness*, philosopher Ted Honderich introduces yet another exotic view, which he calls 'Radical Externalism' or 'Consciousness as Existence'.

Honderich acknowledges the fact of psychoneural intimacy: conscious or mental events are in some kind of necessary connection with neural events. This intimacy has often been taken as a sign of identity, leading to mind–brain identity theories which claim that conscious events *are* neural ones. However, for Honderich and many others consciousness doesn't seem to be that sort of physical thing. They feel that mind–brain identity theories leave out something essential, namely the reality of our mental lives, the most immediate of all the facts we know.

One option for the identity theorist is to admit that it is not possible to reduce consciousness to neural processes as conceived by contemporary neuroscience, but to argue that consciousness consists in some as yet unknown physical process, which future neuroscience may be able to describe. Honderich rejects this option as 'much too strange and adventurous'.

He thinks that conscious events are more than neural, but is unwilling to account for this by postulating any new feature of the world. Thus the non-neural properties of consciousness must somehow be explained in terms of concepts we already possess. But how? He suggests that the key challenge is to find a different way of thinking about familiar ideas. There is what we already know to be inside our heads (the stuff that neuroscience describes) and what is outside our heads (the subject matter of both perceptual experience and science). Because it

seems obvious to Honderich that consciousness has features that don't belong inside heads, he is left with only one option, namely to locate the non-neural properties of consciousness *outside* our heads. And this has to be done without adding anything outrageous to the ontology we already have. The outcome of reasoning within these constraints is what he calls 'Radical Externalism' or 'Consciousness as Existence'.

The starting point of Honderich's new theory is the proposal 'my perceptual consciousness now consists in the existence of a world'. But what does 'world' mean here? He notes that in everyday usage 'world' is a unitary concept. However there are secondary usages involving three worlds that are relatively independent of minds or subjects — i.e. the 'physical world' which is spatio-temporal and has perceived properties or is spatio-temporal and is in nomic (i.e. lawlike) connection with things that have perceived properties; the 'objective world' which has in it things perceivable by more than one person and also exists independently of perception; finally, there is the 'world of things in current and anticipated science'. But Honderich recommends that an adequate account of consciousness requires yet another conception of world. This is the world that *is* my perceptual consciousness. His theory relies on understanding relations between these several conceptions, especially those involving his newly proposed world.

The difference between 'my world of perceptual consciousness' and the mind-independent worlds consists in the former depending on me and my neural events. Further, it is private and does not exist unperceived. However, there is also a large similarity between the world of perceptual consciousness and the perceived part of the physical world. Both are spatial, temporal and contain objects like chairs. A further similarity is that both depend upon neural events and upon yet another world, 'the world-in-itself'. The similarity between the world of perceptual consciousness and the perceived part of the physical world is crucially important for Honderich's theory, for it suggests that 'my world of perceptual consciousness' is a world 'out there' and cannot be regarded as simply an inner mental world of thoughts and feelings. In somewhat Berkeleyan fashion, Honderich states that worlds of perceptual consciousness are the only worlds that are not theoretical. The idea, in a nutshell, is that for something to be conscious is for a world to exist. Thus Honderich's is a view of perceptual consciousness *as existence*, or of existence as perceptual consciousness.

Honderich's strategy is to contemplate consciousness directly. For him the fundamental issue about consciousness is its seeming mystery. With his reconceptualization he hopes to shed light on it, believing that trying to dissipate the mystery is better than recoiling from it. This attitude makes his book stand out from the vast current literature on consciousness. There would never have been a quantum theory if physicists had not appreciated the mysteriousness of certain experimental results, and gone on to seek for new ideas to explain them. Analogously, I think there will never be a theory of consciousness if researchers do not first appreciate and acknowledge how very puzzling it is. Indeed, there is no need for a theory if there is nothing to explain! In this respect Honderich is on

the right track. Unlike many of his physicalist colleagues, he does not sweep the mystery of consciousness under the carpet. It is right there throughout, and its presence gives much force and beauty to the book.

However, one can also wonder about Honderich's strategy, especially his view that no new empirical discoveries are needed to account for consciousness. If this were so, it would be unlike almost anything hitherto encountered by the natural sciences. For if one looks at the history of science one can frequently see how new entities and processes have been proposed to account for puzzling phenomena, and how later developments have either confirmed or questioned the existence of such entities. Atomic theory is perhaps the most strikingly successful example of such a proposal. So I doubt whether Honderich is correct in this respect. I suspect that a deeper understanding of the physiology of the brain (e.g. one using the rich conceptual and empirical resources of quantum physics) is likely to significantly increase our understanding of the relation of matter with consciousness. However, this need not undermine the value of Honderich's approach, for any putative new 'quantum approach' to consciousness would need to answer the profound questions that he so elegantly poses. It is not likely that neuroscience, even when enriched by some deeper theory, could directly tackle the mystery of consciousness. It needs to be complemented with a more philosophical theory, and here Honderich's view could play a key role. So I agree that reconceptualizing what we already know, in the way Honderich proposes, may be a strategy crucial to understanding consciousness. But reconceptualization of the old and search for the new should surely go hand in hand, mutually constraining and complementing each other as we strive toward a better theory.

In conclusion, I think Honderich's approach both new and very interesting. It deserves the attention of both those who view conscious experience as deeply puzzling and those who see no mystery.

### BOOKS RECEIVED

*Mention here neither implies nor precludes subsequent review.*

- Fried, Eric Wolf, *Inwardness and Morality* (Radopi, 2005)  
 Garry, Ron, *Wisdom Nectar: Dudjom Rinpoche's Heart Advice* (Snow Lion, 2005)  
 Jamgön Kongtrul, *The Treasury of Knowledge: Systems of Buddhist Tantra* (Snow Lion, 2005)