Book Reviews

Christina E. Erneling and David Martel Johnson (eds.)
The Mind as a Scientific Object: Between Brain and Culture
New York: Oxford University Press. 2005. xiii + 549 pp.
ISBN 0-19-513932-1; 0-19-513933-X (pbk.)

Reviewed by Valerie Gray Hardcastle
Science and Technology Studies, Virginia Tech.

The Mind as a Scientific Object is a book with a message. At first, I found that fact off-putting in an edited volume. I wanted at least the illusion of being offered a balanced treatment of the issues. Of course most edited volumes aren't balanced, but this volume wears its prejudice on its sleeve — literally. This book is designed to argue that, along with neuroscience, cultural accounts of mind are of fundamental importance to cognitive science and to the very ontology of mind. Not all chapters argue for this point — indeed many do not — but the editors make very clear in their commentary that this is the ultimate conclusion they are after.

It's a huge book with twenty-nine chapters, almost too many for a single volume. Obviously, not all are equally good. For the most part, I found the individual contributions adequate, with a few clunkers that should have been weeded out at an earlier stage. They mainly argue for points or perspectives already found in the literature, so there are very few truly new ideas here. (One exception to this generalization might be Johnson's own contribution, 'Mind, Brain, and the Upper Paleolithic', in which he points out that evolutionary psychological accounts of human development overlook the possibility that our brains evolved to something approximating their present form long before our ancestors started exhibiting any markers of complex thought. As a new and different just-so-story, I found this essay to be fresh and inventive.)

The book is divided into seven sections, each of which is designed to answer specific questions geared toward making the editors' ultimate point: that we had better pay more attention to cultural accounts of mind if cognitive science is to have a prayer of making progress. What I like about their selections and ultimate point is that they are pushing us to re-examine our basic ontology of the mind. I

agree with them that we need to do this; I also agree that this project is important and often gets insufficient emphasis.

However, this is a book by philosophers for philosophers, which means one often gets cartoon accounts of science. I know of few bench researchers who would hold the strong views that the editors claim cognitive science asserts about the mind — few who would disagree with the claim that a cognitive theory is not the whole story. At the same time, I know many philosophers who think that scientists hold such unidimensional perspectives. This book is for them.

The first section asks, 'Where are we [at] present and how did we get there?' It underscores general pessimism about progress in cognitive science as a purely scientific discipline and diagnoses this failure as due to psychology's limitations. The remedy on offer is to see that mind is as much a cultural object — and therefore an object of humanistic inquiry — as it is a scientific one. I am generally not sympathetic to these sorts of argument, for I find that they often (always?) misconstrue the task of science. After all, science does not preclude other disciplines from also investigating things, using whatever methodologies they wish.

Section two asks, 'Is the study of mind continuous with the rest of science?' The short answer is 'no, it is not'. If anything, cognitive science is a 'practical' discipline instead of a theoretical one. That is, it is designed to help us help others and not to tell us how things really are. In particular, arguments about dualism (regardless of whether you believe them) tell us that the mind is somehow special from a scientific point of view. The editors want to conclude that that specialness reflects the mind's non-scientific aspects.

Section three queries whether eliminative materialism is 'sound or mistaken'. There is unanimous agreement that it is mistaken. If there is debate at all, it is over where the mistake lies. For those interested in such debate, this section might be worth perusing. For the rest of us, the topic is too worn to add much to the overall theme of the book, or to philosophy of mind in general, for that matter.

Section four treads old ground as well, asking whether '"mind" is just another name for brain and what the brain does'. In other words, should we expect a reductionistic account of the mind out of cognitive science? This was the weakest part of the book, with the first three contributions simply summarizing well-known arguments and positions. The final essay, 'Gall's Legacy Revisited: Decomposition and Localization in Cognitive Neuroscience,' by Zawidski and Bechtel, does give a nice case study in attention that shows by vivid illustration the complexity of reductionism in the actual practice of science, and so stands apart from the previous three chapters.

The final three sections really provide the meat of the book. Number five inquires whether evolution provides 'a key to the scientific study of mind?' It answers with an emphatic yes. Each of the chapters takes issue with the idea that humans are somehow special animals and each emphasizes the deep connections between culture and nature in human development. That is, culture is part of nature — certainly part of our nature and probably that of other creatures too — and a complete understanding of mind is going to be sensitive to this fact.

Section six asks 'Is the mind a cultural entity?', again answered in the affirmative. Each of the chapters, by way of commentary on other authors, argues that traditional cognitive science is much too narrow in how it approaches the study of mind. We need to 'go literary' (to borrow Jerome Bruner's apt phrase) in order to better appreciate culture's connections to mind, not just by reading more literature but by genuinely integrating literary concerns into the fields of psychology, linguistics and biology.

The final part, 'Rationality: cultural or natural?' illustrates the point of section six via studies in rationality. The authors argue that the social side of rationality is vitally important in understanding how humans reason. Cognitive science needs to move beyond considering how individuals reason to appreciate this idea. As a result (and this truly is the point of the book), we must reconceptualize our notion of the mind; we must view it as an heterogeneous entity, one that is both cultural and biological. The mind is essentially both, and our science needs to reflect that fact.

This book contains some interesting ideas and certainly its overarching theme is well worth considering. However, I don't believe that it needed twenty-nine chapters to make its point, so I recommend picking and choosing. You can do so without losing too much, especially if you already know a little about recent debates in the philosophy of cognitive science. I would focus my attention especially on the final three sections, using the others as background reference material. If you do that, then reading (half) the book is worthwhile.

Igor Aleksander
The World in My Mind, My Mind in the World
Exeter: Imprint Academic, 2005, 196 pp. $34.90 / £17.95
ISBN 1-84540-021-6 (hbk)

Reviewed by Alwyn Scott
University of Arizona

In the fall of 1951, an MIT student drama group put on *RUR (Rossum's Universal Robots)* by the Czech writer Karel Capek — a play from the 1920s in which the term 'robot' was coined from the Slavic word for 'worker'. Appropriately, mathematics professor Norbert Wiener introduced the action with some ideas from his recent books on cybernetics and the looming dangers of the information age, after which he signaled for the entrance of a rudimentary robot — a little wagon with an electronically controlled motor on each wheel and an electric eye so that it moved in the direction of light. Dramatically pulling a flashlight from his pocket, Wiener shone it at the curtain as some reckless robot wrangler kicked the poor little cart out onto the stage. Unfortunately, one of the wheel motors was damaged (perhaps by the kick?) so all the thing could do was turn around and around in circles, ignoring the flashlight. The house was vastly entertained, but was it nervous laughter?

Fifty-four years later, Igor Aleksander tells us that the concept of a conscious robot is not so crazy after all. A professor emeritus of electrical engineering at Imperial College, London and an established author of books on the brain for both specialist and general readers, Aleksander rejects philosophical claims that machines cannot be conscious. His strategy is to introduce five axioms that are essential for an entity — human, animal or machine — to have consciousness: first, a feeling of being both part of and separate from an 'out there' world; second, having a perception of the world that mingles with feelings of past experience; third, having selective and purposeful experiences of the world; fourth, constantly thinking ahead and trying to decide what to do next; finally, having feelings, emotions and moods that help to make decisions.

He uses the axioms as a basis for analysing the need for sleep and the functions of dreaming, relating his discussion to Sigmund Freud's concept of the unconscious. Employing somewhat more technical ideas sketched in an appendix, he then provides concrete examples of his various theories — a feature that will be appreciated by those (i.e. engineers) who must design things that actually work. This discussion is followed by analysis of perception, leading to demolition of various philosophical arguments about the illusoriness of perception.

The most interesting chapter, in my view, discusses free will. Following an historical review of this age-old problem, Aleksander presents a specific example of a machine making a free choice. As such a demonstration is missing from most philosophical treatments of free will, Aleksander's discussion is refreshing and should interest many readers of this journal. The analysis of free will is followed by critical examination of the widely discussed 'explanatory gap' (Levine's term), which underlies David Chalmers' 'hard problem'. Here, Aleksander's axiomatic approach eliminates the 'logically possible zombie' — upon which several of Chalmers' arguments depend. Whereas Chalmers assumes an explanatory gap to exist, in other words, Aleksander's account has no place for one. Finally, Aleksander responds to arguments against machine consciousness, comparing generalizations from the fact that laptop computers aren't conscious to those against heavier-than-air flight, based on the fact that bicycles can't fly.

Although this book has much to offer current discussions on the nature of consciousness and I enjoyed reading it, there are two aspects that could be improved upon. First, while he mentions Donald Hebb's *Organization of Behavior*, it is not noted how many of his ideas stem from this classic work. For example, his 'module' and 'state-space trajectory' were introduced by Hebb as a 'cell assembly' and a 'phase sequence'. Furthermore, in discussing the persistent nature of activity in closed causal loops, Aleksander cites Hebb's book only for the comment: 'In the old days this [phenomenon] used to be called a reverberatory circuit.' Compare this with what Hebb actually wrote in 1949:

> Any frequently repeated, particular stimulation will lead to the slow development of a 'cell-assembly,' a diffuse structure comprising cells . . . capable of acting briefly as a closed system, delivering facilitation to other such systems and usually having a specific motor facilitation. A series of such events constitutes a 'phase

> sequence' — the thought process. Each assembly may be aroused by a preceding assembly, by a sensory event, or — normally — by both. The central facilitation from one of these activities on the next is the prototype of 'attention.'

Hebb described the cell assembly as being like a 'three-dimensional fish net' spread out over the brain; thus he is writing about a closed causal *network*, not a loop. I raise this point because Hebb's seminal work has been ignored over the past half century by researchers who have come up with essentially the same ideas. Thus many *JCS* readers will be interested in Hebb's mature formulation of his cell assembly theory (Hebb, 1980) and in recent neurologically reasonable computer simulations by Anders Lansner and his colleagues at Stockholm's Royal Institute of Technology (Fransén & Lansner, 1995; 1998; Fransén *et al.,* 1980; Lansner, 2005).

Second, Aleksander downplays the importance of hierarchical organization in biological brains, which are inherent in Hebb's theory (Hebb, 1980). Importantly, the hierarchical structures of biological organisms and their cognitive systems open the possibility of closed causal nets that span several levels of both hierarchical systems (biological and cognitive), leading to natural phenomena so intricate that physical science has not yet begun to describe them — either theoretically or by experimental observations.

Perhaps something like this was passing through Norbert Wiener's mind at the close of his introductory remarks to Capek's play, as he told his audience of young technologists — over a half century ago — that to deal with the social and spiritual challenges of the oncoming information age:

> *Either engineers must become poets or poets must become engineers.*

References

Fransén, E. and Lansner, A. (1995), 'Low spiking rates in a population of mutually exciting pyramidal cells', *Network*, **6**, pp. 271–88.

Fransén, E. and Lansner, A. (1998), 'A model of cortical associative memory based on a horizontal network of connected columns', *Network*, **9**, pp. 235–64.

Fransén, E., Lansner, A. and Liljenström, H.(1993), 'A model of cortical memory based on Hebbian cell assemblies', in *Computation and Neural Systems*, ed. F.H. Eeckman and J.M. Bower (Boston, MA: Kluwer).

Hebb, D.O. (1980), *Essay on Mind* (Hillsdale, NJ: Lawrence Erlbaum Associates).

Lansner, A. (2005), 'Cell assemblies', in *Encyclopedia of Nonlinear Science*, ed. A.C. Scott (New York: Tayor & Francis).

Stefano Franchi and Güven Güzeldere (eds).
Mechanical Bodies, Computational Minds
Cambridge, MA: MIT Press, 2005, 538 pp., £29.95
ISBN 0262562065

Reviewed by David W. Salt
University of Derby

Colleagues in my university Computing department often refer to business people as 'arm waving'. Indeed some even regard systems analysis as an 'arm

waving' subject. The implication is clear — real computing is an engineering discipline and everything that falls outside a very narrow band of technical work is not worthy of consideration. Taking a yet smaller slice of the already narrow band of computing reveals the even more constrained field of Artificial Intelligence (AI). A measure of how tightly the borders of this self-defined sub-discipline of computing are patrolled can be found in the web pages of the *JCS* where Pat Hayes comments 'it's a pity that someone who apparently has a post in a school of computing knows so little about the field he professes'. The remark refers to my heretical view that computer scientists working in AI often mistake metaphor for reality. I wear this remark as a badge of honour and often quote it to my students.

All this is by way of preamble to what is an excellent and welcome collection of papers, which seeks to provide 'a forum for intellectual exchange between the artificial intelligence (AI) community and scholars in the traditional humanities and social sciences'(p.1). Although many other publications have sought to explore the inter-disciplinary nature of AI research this book is easily the best that I have come across. The problem it faces is that it seems there will always be a hard core of AI researchers who insist on treating anything even marginally outside of traditional computing as 'arm waving'. The editors and contributors are well aware of this problem and address it extensively within the book, but that does not mean it will just go away.

Usually collections of this type have several below par contributions but here everyone pulls their weight and it was a pleasure to read so many cogent and well-constructed arguments. I often found myself disagreeing with some of the points being made but none of the articles can be accused of being of poor standard.

The first chapter by the editors Stefano Franchi and Güven Güzeldere runs to over 100 pages, plus an additional 26 pages of references, and is itself worth the price of the book. I still make reference to Güzeldere's excellent 10 year-old overview of writings on consciousness (in *JCS* vol. 2). Franchi and he give a similar account here, a meticulous summary which sets the historical detail of AI in context and explores all the significant developments and arguments that have dominated the field over the last sixty years. It also examines the precursors of the discipline, as do several other contributors. Overlaid on this review is a critique which broadly argues that engineering alone is insufficient as a basis for building AI; that the key to success is likely to be in embodying intelligence and that a likely way forward lies in a revival of cybernetics. More specifically the editors argue that the convergence of AI and the currently out-of-fashion field of cybernetics could lead us to conclude that we should be seeking to produce cybernetic organisms. They report that there is currently much more interest from the humanities in cyborgs than there ever was in AI. If cyborgs 'solve' the problem of embodiment and humanists are as interested as the AI researchers, there may be a real opportunity for new cross-disciplinary initiatives. I cannot find fault with this section of the book and only regret that they did not find space to explore the related topic of transhumanism.

Elsewhere in the book you can find a broad range of related topics. Everything from post-modern and semiotic analyses to logic and its uses and abuses; from the role of socialisation in creating intelligence to amusing conversations between so-called intelligent programs; from the heavy-hitters of AI to a whole range of contributions from the humanities. It is impossible to do justice to all contributions in a review of this kind so I have picked out just a few.

In the heavy-hitter category, there is a useful exchange between Dreyfus and Dennett which starts as a debate on the significance of Deep Blue's win over Kasparov and moves on to the abilities of 'Cog' (MIT's long-running project to embody AI in a robot), neatly underpinning some of Franchi and Guzeldere's key conclusions. Fred Dretske gives a careful analysis of what it means to have mentality. In 'proving' that neither computers nor people are really performing addition when they add numbers together, he elegantly demonstrates that consciousness is just as hard to demonstrate in humans as it is in machines. His argument has something in common with Searle's Chinese Room thought experiment, whilst his conclusions veer towards the systems and emergence views found in Dennett. (I was also very taken with Dretske's use of 'fleet' as a collective nouns for gerbils, as in 'a fleet of gerbils'.) Douglas Hofstadter argues that logic alone is too brittle for AI systems, whilst perhaps paradoxically concluding that a mathematical formulization of 'seeing as' may be the missing link in AI.

Alison Adam's 'Knowing Subjects: AI from Feminist Philosophy' was the paper that provoked me most. I also found it the most insightful and useful in pinpointing what's missing from many AI projects. In other words I largely agreed with her conclusions though not necessarily with her viewpoint. She argues that the epistemology of AI is essentially gendered. Two key pieces of evidence here are:

(1) the agent who perceives in AI is always described in objective terms rather than subjectively in the way that humans experience events;
(2) women hold knowledge within their bodies that is not represented in AI systems.

From this evidence she concludes that AI epistemology is mainly masculine and hence misses some key aspects of intelligence. 'CYC', Douglas Lenat's gigantic rule-based project which has been in production for many years, is cited as one example of an AI system which has had poor success precisely because it is a paradigm of masculine epistemology. Intuitively I think that Adam is right — that subjectivity and body knowledge are missing in most extant AI systems and that this needs to be urgently addressed. Moving to an embodied form of AI or cybernetic organism ought to correct some of these imbalances. Where I disagree with her is in her view that the 'normal' AI epistemology is male and that subjective / body knowledge is largely female. Every intelligent being brings subjectivity and body knowledge to their experience of the world. The fact that some men choose to deny this does not mean that these attributes belong to women alone. Nevertheless her apparent denial of embodiment to men does not detract from the fact that I think this is the one paper that really hits the nail on the head when

it comes to recognising what is missing in Good Old-Fashioned AI (GOFAI) and what needs to be incorporated into future cybernetic organisms.

This is a book that deserves to be in the library of everyone working in computing as an AI expert, and everyone in other disciplines who wants to be involved in AI or to understand its history and why it is so important. It is the best written and best argued case for cross disciplinary work that I have ever read and my only concern is that it may not be read by the unconverted. The AI 'establishment' may dismiss it as mere 'arm waving', which would be very unfortunate.

Eva Jablonka and Marion J. Lamb
Evolution in Four Dimensions: Genetic, Epigenetic, Behavioural and Symbolic Variation on the History of Life
Cambridge, MA and London: The MIT Press, 2005. £22.95 (hbk)
462 + x pp. ISBN 0-262-10107-6.

Reviewed by Chris Nunn

Less than a generation has passed since Richard Dawkins and Stephen Jay Gould fought their epic battles, yet already they seem one with Hector and Achilles; mighty warriors for sure, but ones whose deeds resounded in a distant, narrow world. Evolutionary studies have blossomed out of all recognition recently, thanks in large part to the discovery that deciphering genomes raises far more questions than it answers. You could not find better guides to the new vistas than Jablonka, a philosopher specialising in the history of ideas, and Lamb who is a biologist.

They show that DNA-based selfish genes, often considered the be-all and end-all of heredity, are only one among four relevant systems, albeit the most fundamental. Traditional dogma has it that natural selection acting on consequences of entirely random DNA mutation is wholly responsible for evolution. Even at this bottom level, however, organisms can sometimes accelerate rates of mutation when they need to meet an environmental challenge. More remarkable still, it seems that they may selectively accelerate mutation rates in the particular genes that are likely to throw up an answer to whatever challenge it may be. The dice, it seems, can be loaded in a genome's favour.

Next up are 'epigenetic' factors. There is a range of these (at least four), which have the common property of allowing transmission to its descendents of variations acquired by a cell. Their most obvious function is to ensure that, in multicellular creatures, cell lines breed true; to ensure, in other words, that kidney cells give rise to kidney cells and not skin cells, for example, or neurons. All our cells contain the same genome, so the system ensuring that only the right bits of it are activated at the right times and places has to be powerful and reliable. Epigenetic factors not only control the fate of cells within an individual, but can also be transmitted down the generations. This commonly happens in simple organisms and plants but there's also strong evidence for its importance in animals, including us. In relation to ourselves, most of the evidence has to do with

hereditary diseases but, if epigenetic factors can be responsible for these, they are likely to play more constructive roles too.

The third system involves naïve individuals copying aspects of the behaviour of experienced ones. A fairly advanced brain is needed for this, but what's involved is not always obvious. For example, it has been shown that rabbit pups are predisposed to eat whatever was in their mother's diet as a consequence of pre-natal, as well as post-natal, influences. Apparently, even as foetuses, they are somehow able to access information about what their mother regards as nutritious food, which they put to good use later.

Finally there are transmissible symbols. Some advanced animals make limited use of them, but they show themselves to full effect only in our own languages and cultures. Jablonka and Lamb make the interesting point that this last 'heredity' system is like the DNA one in that it is modular, has dedicated copying system(s), can transmit latent information and has unlimited variation. The two intermediate systems (epigenetic and behaviour copying) generally lack these qualities.

Having described and analysed their four systems, the authors then have to put the evolutionary Humpty Dumpty back together again. It is here that they take their most radical step for they point out that, with the exception of DNA, all the systems involve transmission of *acquired* information and are thus Lamarckian. How can such a seemingly heretical observation be reconciled with orthodox Darwinism? They explain matters with great lucidity and elegance, aided by nice illustrations (drawn by Anna Zeligowski) plus Socratic dialogues between the authors and an *alter ego* playing the part of devil's advocate. What it all boils down to is that the three non-DNA systems put constraints on the probable direction of any Darwinian evolution. The 'Watchmaker' is still blind, but his hand is guided towards a greater chance of finding materials needed for his craft. The guide is not any sort of supernatural final cause, but rather the outcome of evolution and selection occurring in the Lamarckian systems, which are in turn affected by whatever happens in the DNA one. It's a beautiful and convincing picture, though inordinately complex.

How complete is it? Well, only the outlines are available so far. The complexity is such that it will be many years before details of even the most important feed-back loops become known for sure. A piece is missing, too, from the purely Darwinian parts of the picture. Jablonka and Lamb paint this in a rather traditional way as dependent on competition — survival of the fittest and all that. They make occasional nods in other directions, but these never amount to much. However there's lots of evidence that, while competition is essential for the 'fine tuning', cooperation is also vital especially to major evolutionary change (see for instance Frank Ryan's accessible account in his book *Darwin's Blind Spot*).

Then the discussion of memes — the gene-like 'units' of culture proposed by Dawkins thirty years ago — comes across as less than cogent despite their possible relevance to the 'symbolic' system. The concept remains controversial, with supporters and detractors fairly evenly balanced. Detractors usually base objections to it on the lack of any good definition of a meme, arguing that no such

definition is possible. Jablonka and Lamb belong in the doubter's camp, but their own main objection is that memes are not transmitted from person to person but are 'reconstructed' within each individual. However, exactly the same could be said of DNA, which is synthesised on the basis of templates provided by parents. Of course the reconstruction is far more faithful to the template in the case of a gene than in that of any hypothetical meme, but that's a difference of degree not of kind. Further, the fact that they regard symbolic systems as 'modular' suggests a basis for satisfactorily defining memes — i.e. as transferable 'modules'. Perhaps they should consider switching camps.

Whatever the problems of detail the overall sweep of the argument is compelling, based on a very broad range of fascinating evidence. The final chapter describes some implications. On the plus side, better understanding of epigenetic systems in particular holds great promise for improved medical treatment. On the other hand, evolved systems are so complex that they can never be recovered once destroyed. Jurassic Park is strictly a fantasy. In one lucid package, this book tells us a huge amount about ourselves and the world we inhabit. It deserves a place on everyone's 'must read' list.

Jaegwon Kim
Physicalism, or Something Near Enough
Princeton University Press, 2005. xiii + 186 pp. $26.95 / £17.50
ISBN 0691113750 (hbk)

Reviewed by Dimitris Platchias
University of Glasgow

In his new book, Jaegwon Kim argues that, although physicalism may not be the whole truth about the mind–body problem, it is the truth near enough. Kim offers an overall view of the philosophical landscape of the problem thereby determining the kind of physicalism to which we can lay claim. The book is for the most part an attempt to defend views on issues of mental causation and consciousness that appear in his earlier work *Mind in a Physical World* (1998). There, Kim argued against currently popular middle ground positions in the philosophy of mind. According to him, the many different versions of non-reductive physicalism founder on the problem of mental causation; therefore to be both a physicalist and maintain the causal efficacy of the mental, one must be a reductionist.

Now and then Kim uses the notion of supervenience. One of the advantages of this relation of dependency between the mental and the physical (crudely construed: no mental difference without a physical difference) is its clear implication that the mental domain is anchored in the physical domain without at the same time implying physical reductionism. Supervenience is generally construed as an asymmetrical relation (as opposed to the identity relation): the mental is determined by or dependent on the physical, not the other way round. It appears though, that the supervenience relation states only a pattern of

covariance between mental and physical properties and does not explain those correlations (why, for instance, C-fibre firing correlates with the sensation of pain). The postulation of empty necessities where we do not understand how the connections are necessary can certainly mislead. Moreover supervenience is compatible with numerous doctrines in the philosophy of mind that are themselves incompatible, such as epiphenomenalism and type-physicalism. Kim (1998; 2005) acknowledges these problems and argues that supervenience alone is not sufficient for physicalism. It is however, important and includes a claim of existential dependence of the mental on the physical.

This new book contains six chapters/independent essays, each engagingly written and a model of clarity. In Ch. 1, Kim formulates two mind–body problems, namely mental causation and consciousness. According to him, these two 'world-knots' are interlocked and so the problem of mental causation is solvable only if mentality is physically reducible, though phenomenal consciousness resists physical reduction. In Ch. 2, he presents a more detailed formulation of his supervenience/exclusion argument and addresses two well-known objections to it, namely that causal overdetermination is still an option and the generalisation argument (Noordhof, 1999; Block, 2003).

The non-reductive physicalist holds that a mind–body supervenience relation exists, that the physical is irreducible to the mental and that the mental is causally efficacious. According to the physicalist though, the physical domain is causally closed, therefore (and this is very crudely how Kim's supervenience/exclusion argument goes) if mind–body supervenience fails, then mental causation is impossible because that would amount to violation of the principle of causal closure of the physical. If mind–body supervenience holds, then again the mental is causally inert because physical events are not causally or nomologically overdetermined. Causal overdetermination (a case where a physical property P_1 and a mental property M_1 are *genuine* causes of M_2) is untenable. The physicalist, then, must choose between causal impotence and reduction. According to Kim, mind–body supervenience does hold, so does the *Exclusion* principle,[1] and the physical domain is causally closed: 'the conclusion is that causally efficacious mental phenomena must be reducible to physical ones' (p.153).

Kim argues that all causal powers are contributed by physical properties. But, if one holds the *Exclusion* principle and the principle of the causal closure of the physical, then arguments analogous to Kim's show that not only is there no mental causation but no physiological causation, no molecular causation but only bottom level physical causation (if there *is* a bottom level of physics — otherwise no causation at all). It appears that the supervenience argument *generalises* beyond mind–body causation: since properties at upper level are supervenient on lower level properties, causation at any level gives way to causation at the next

[1] This principle states that no single event can have more than one sufficient cause occurring at any given time unless it is a case of causal ovedetermination. Kim is very convincing in defending his view that rules out causal overdetermination as an option. The *Exclusion* principle does not favour the mental or the physical cause but the principle of the causal closure of the physical excludes the mental cause, enabling the physical cause to prevail.

lower level. Kim's less than fully satisfactory reply is that (only) the fundamental level of microphysics is causally closed and we need this in order to have the required causal premise available. Moreover, he questions the plausibility of assumptions such as the irreducibility of the biological level to the physicochemical, stating that 'the reduction option must be ruled out for purely physical levels, including microphysical levels, and it is far from obvious that this can be done' (p. 69).

In Ch. 3, he formulates an argument against dualism-interactionism to the effect that it offers no help with mental causation. In the next two Chapters, he discusses two options for closing the explanatory gap (Levine, 1983), namely reduction and reductive explanation, and opposes arguments for type-physicalism, (one by McLaughlin and Hill and one by Block and Stalnaker), to the effect that psychoneural identities do not explain psychoneural correlations, 'identities do not seem capable of generating explanations on their own; the best they can do is to "transfer" explanations that have already been completed … from one *description* of a phenomenon to *another description* of the same phenomenon'(p.146). His illuminating discussion on whether the identification of phenomenal with neurophysiological properties can be supported empirically with inference to the best explanation is one of the most interesting parts of the book.

In the last Chapter, Kim arrives at a position that he calls 'conditional physical reductionism'. On this account, if the mental is causally efficacious it must be physical and the reductive option is functional reduction (against the mainstream physicalist view — due to multiple realisation — and tantamount in some respects to a restricted kind of type-physicalism). He concludes that, whereas intentional properties such as beliefs and desires are functionally reducible, qualia (phenomenal consciousness) are not so reducible. We are not however, losing sensory experiences altogether. Certain relational properties of qualia, in particular their similarities and differences, are in principle functionally reducible. What can't be saved are the intrinsic qualities of our sensory experiences (the fact, as Kim says, that yellow looks like this, that ammonia smells like that and so on): 'intrinsic qualities of qualia are irreducible, and hence causally impotent. They stay outside the physical domain…' (p.173). But consciousness in the most interesting sense *is* phenomenal consciousness and what makes a state phenomenally conscious is that there is something it is like to be in that state (Nagel, 1974; Block, 1995; Chalmers, 1996). Kim however concludes by saying that losing phenomenal consciousness isn't losing much. At least he admits that phenomenal consciousness is not a fiction of bad philosophy and that it should be explained and not explained away. Is there any other option for the physicalist? Representationalism, the currently favourite approach for naturalising the mind is left undiscussed.

References

Block, Ned (1995), 'On a confusion about a function of consciousness', *Behavioral and Brain Sciences*, **18**, pp. 227–47.

Block, Ned (2003), 'Do causal powers drain away?', *Philosophy and Phenomenological Research,* **67**, pp. 133–50.
Chalmers, David (1996), *The Conscious Mind: In Search of a Fundamental Theory* (New York: Oxford University Press).
Kim, Jaegwon (1998), *Mind in a Physical World* (Cambridge, MA: MIT Press).
Levine, Joseph (1983), 'Materialism and qualia: The explanatory gap', *Pacific Philosophical Quarterly*, **64**, pp. 354–61.
Nagel, Thomas (1974), 'What is it like to be a bat?', *Philosophical Review*, **83**, pp. 435–50.
Noordhof, Paul (1999), 'Micro-based properties and the supervenience argument', *Proceedings of the Aristotelian Society*, **99**, pp. 109–14.

Michael S. Gazzaniga
The Ethical Brain
New York: Dana Press, 2005. 225 pp., $25, ~ £14
ISBN 1932594019 (hbk)

Reviewed by Stephan Schleim

Over the past five years, the emerging field of neuroethics has received ever more public attention. Given the potential ramifications of neuroscientific research, the trend is not surprising and is socially necessary. Because Gazzaniga is a member of the US President's Council on Bioethics, his opinions deserve special consideration. This new book summarizes his views on ethical issues surrounding the beginning and end of human life, brain-enhancement, free will and responsibility, and the nature of moral beliefs.

Unfortunately, though, he does not devote much attention to the 'ethics' in 'neuroethics' and this imbalance shapes the whole book. His bias may ultimately derive from an omnipresent, naive optimism about scientific research in general and the self-regulation of scientists in particular. The 'magic hand' of Adam Smith's market applies equally, he seems to think, to science. For example, discussing the possibility of brain-enhancement through genetic modification of the embryo, he invokes the example of sex selection; some biomedical companies offer their services for the sake of 'family balancing' only from the second child onwards, suggesting to Gazzaniga that 'those *closest* to new technologies immediately respond with good sense' such that 'the greater social good prevails' (p. 51). For him, nothing using our 'very human skill,' i.e. the human brain, deserves to be called 'dehumanizing' (p. 53), an idea that mass murderers probably would welcome. But his analogy is just as unapt as his other example, the atomic bomb: 'Sure we humans built it, but we humans are dead set on never using it again' (p. 53) — how realistic this is in the age of more than 30,000 nuclear warheads distributed over at least five countries, the reader may decide himself.

In the hotly debated issue of pharmacological augmentation, Gazzaniga distinguishes between physical and mental enhancement. Legalizing the former, for example doping in sports, would have implications in a public context and lead to a pharmacological 'arms race' that neutralized 'the whole logic of competition' (p. 57). Further, this kind of manipulation would not only affect an individual's self-image and that of others (p. 64), but also hold the risk of 'deleterious

long-term consequences to the physical nervous system' (p. 62). So far, it is clear why Gazzaniga rejects physical enhancement as illicit cheating. The argument becomes less clear, however, when it comes to the case of *mental* enhancement. To say nothing of unknown risks that may arise with the long-term use of drugs like Ritalin® (methylphenidate), it remains an open question why their use, if legalized, should have less social impact than physical enhancers as they allegedly increased performance in SAT entrance examinations by more than a hundred points (p. 72). Surely we can expect a large potential of *coercion* arising from the use and abuse of these drugs (Farah *et al*., 2004; Chatterjee, 2004; Farah, 2005). Further, considering that the worldwide market volume of so-called 'lifestyle drugs' was worth US$ 20 billion in 2002 and is estimated to increase to US$ 29 billion by 2007 (Atkinson, 2002; Flower, 2004), it is unlikely that there would be no arms race for shares in the market for cognitive enhancement. Another of his arguments in favour of mental enhancement is that, 'in some way we were cheated by Mother Nature if we didn't get the superior memory system, so for us to cheat her back through our own inventiveness seems like a smart thing to do' (p. 73). This immediately raises the question why the racing cyclist who got a cardiovascular and nervous system inferior to Lance Armstrong's, say, should not be allowed 'to cheat back' in the same manner — a case that Gazzaniga explicitly rejects.

He has a genuine point, however, when he addresses the way the public usually deals with putative consequences of research. Far too often, fears are stirred by 'slippery slope' arguments, which Gazzaniga legitimately wants to eliminate from neuroethical discussions (p. xvii). But he errs in that these arguments usually come from self-appointed moralists, not from educated moral philosophers and ethicists. Further, while it indeed makes sense to first consider the *possibility* of future technologies like genetic engineering, instead of jumping to debate all the potential risks, it also hints at a role for lack of public information (a responsibility still too often neglected by scientists) if most people are susceptible to the horror scenarios used by mass media to grab attention. For example, while earlier discussions of stem-cell research were influenced basically by religious provisos on the one hand and exaggerated fantasies about its medical value on the other, they are now becoming constrained by realistic scientific arguments (Pompe *et al.*, 2005). Another benefit of this public ethical controversy has been that many societies have begun to discuss the issue of when the human life begins, which has already had concrete effects on several countries' legislation, as for example the licensing of research on embryos up to 14 days after conception in the UK. Unfortunately Gazzaniga cannot earn credibility for his own approach in this connection; he opts for the eighth week on the basis of a 'gut reaction' (p. 8), self-admittedly contradicting neuroscientific knowledge.

When he gets to questions of free will, personal responsibility and the law, even his most faithful colleagues might begin to demur. For him, the issue of responsibility is a social choice to which neuroscience cannot contribute (pp. 101–2): 'the brain is determined, but the person is free' (p. 99), Gazzaniga writes. At this point, the distinction between *normative* and *descriptive* ethics,

which is wanting throughout the whole book, would have been useful: it is true that neuroscientists cannot *prescribe* the meaning of responsibility. But they can well *describe* whether the condition of a subject is influenced by a pathological condition. For example, lesions to the orbital and medial prefrontal cortex and the amygdala have been shown to lead to impulsive and aggressive behaviour (Grafman *et al*., 1996; Anderson *et al*., 1999; Davidson *et al*., 2000; Blair, 2004). The example of a 40-year-old schoolteacher with an egg-sized tumor in the right lobe of his orbitofrontal cortex who suddenly exhibited uncontrollable paedophilia and general sex-obsession is a well-known case (Burns & Swerdlow, 2003).[2] After surgical removal of the tumor, his criminal behaviour disappeared completely. In such a case, our traditional concept of punishment does not make sense and therefore neuroscientists can, just as generations of psychiatrists did before them, make an important contribution to legal practice. Furthermore, Gazzaniga contradicts himself in saying that 'people are free and therefore responsible for their actions; brains are not responsible' (p. 89), because earlier in the book, in the imaginary case of brain-transplantation, he equated the brain with personhood: 'you are your brain. The neurons interconnecting in its vast network […] — that is you' (p. 31). But if I am free while my brain is not, then obviously I cannot be (just) my brain.

To sum up, *The Ethical Brain* offers a nice overview of current empirical research relevant to the field of neuroethics. To the ethically interested reader, however, it does not have much merit and carries potentially dangerous implications due to the lightheaded way in which it discounts the ethical dimension of the imminent *neurorevolution*, which is likely to have a huge impact on our social world very soon. It is true that ethicists should not generally squelch scientific progress; it is just as true, however, that mankind does not always have to commit mistakes before recognizing them as such.

References

Anderson, S.W., Bechara, A., Damasio, H., Tranel, D. & Damasio, A.R. (1999), 'Impairment of social and moral behavior related to early damage in human prefrontal cortex', *Nature Neuroscience*, **2**, pp. 1032–7.

Atkinson, T. (2002), 'Lifestyle drug market booming', *Nature Medicine*, **8**, pp. 909.

Blair, R.J.R. (2004), 'The roles of orbital frontal cortex in the modulation of antisocial behavior', *Brain and Cognition*, **55**, pp. 198–208.

Burns, J.M., Swerdlow, R.H. (2003), 'Right orbitofrontal tumor with paedophilia symptom and constructional apraxia sign', *Archives of Neurology*, **60**, pp. 437–40.

Chatterjee, A. (2004), 'Cosmetic neurology: The controversy over enhancing movement, mentation, and mood', *Neurology*, **63**, pp. 968–74.

Davidson, R.J., Putnam, K.M. & Larson, C.L. (2000), 'Dysfunction in the neural circuitry of emotion regulaton: A possible prelude to violence', *Science*, **289**, pp. 591–4.

Farah, M.J. (2005), 'Neuroethics: the practical and the philosophical', *Trends in Cognitive Sciences*, **9**, pp. 34–40.

Farah, M.J., Illes, J., Cook-Degan, R., Gardner, H., Kandel, E., King, P., Parens, E., Sahakian, B. & Wolpe, P.R. (2004), 'Neurocognitive enhancement: what can we do and what should we do?', *Nature Reviews Neuroscience*, **5**, pp. 421–5.

[2] For a more public description, see http://www.newscientist.com/article.ns?id=dn2943

Flower, R. (2004), 'Lifestyle drugs: pharmacology and the social agenda', *Trends in Pharmacological Sciences*, **25**, pp. 182–5.
Pompe, S., Bader, M. & Tannert, C. (2005), 'Stem-cell research: the state of the art', *EMBO Reports*, **6**, pp. 297–300.
Grafman, J., Schwab, K., Warden, D., Pridgen, A., Brown, H.R. & Salazar, A.M. (1996), 'Frontal lobe injuries, violence, and aggression: a report of the Vietnam Head Injury Study', *Neurology*, **46**, pp. 1231–8.

Peter J. Ludlow, Yujin Nagasawa & Daniel Stoljar (eds.)
There's Something About Mary: Essays on Phenomenal Consciousness and Frank Jackson's Knowledge Argument
Cambridge, MA: MIT Press, 2004, 480 pp., $35, £22.95
ISBN 0262621894

Reviewed by Shannon Vallor
Santa Clara University

Frank Jackson's 'knowledge argument', offered in his 1982 paper 'Epiphenomenal Qualia', helped to ignite the contemporary philosophical debate over consciousness. Contributing to the argument's impact was its illustration by means of a provocative thought experiment. Though the case of Mary the colour scientist received only a cursory mention in Jackson's original paper, she became the focus of ensuing debate over the merits of the knowledge argument. The story is now familiar: Mary is confined to a black and white room from birth, prevented from ever experiencing colours, all the while receiving a complete education in the neurophysiology of colour perception. The consequent dilemma is clear — what, if anything, would Mary learn upon leaving her room and seeing colours for the first time?

This volume presents an overview of the history of the debate provoked by Mary's case and its implications for physicalism. In addition to Jackson's 1982 and 1986 papers, the volume includes fifteen essays (four previously unpublished) along with a foreword and several postscripts by Jackson. The selections here are solid and well organized. The editors' introduction provides useful historical context for the intuitions behind the knowledge argument. They divide the volume into five sections, each reflecting a different way of responding to Jackson's claims. The first is to challenge the intuition that Mary learns anything at all when she leaves her room; an approach represented here by an essay from Daniel Dennett, followed by a response from Howard Robinson. This section is unsatisfying as the bogey of epiphenomenalism considerably muddies the waters here. More space is devoted to the remaining four types of response, all of which assert that, whatever Mary may learn, it is not of a character to refute physicalism. These responses include the well-known claim that she merely acquires new abilities; the claim that she acquires knowledge by acquaintance rather than factual knowledge; the claim that she achieves a new mode of knowledge of old facts and, finally, the claim that Mary in her room could not have known all the physical facts after all, not even in principle.

Given that Jackson has since withdrawn his support for the knowledge argument, joining many of his earlier critics in regarding the intuitive force of Mary's case as illusory, one might think that his postscripts would comprise the most enlightening part of this volume. Yet this is not the case. Jackson's new stance is premised on a fairly thin representationalist account of perception, and an associated distinction between intensional properties and instantiated ones. This distinction is not as successful as Jackson hopes in 'explaining away the epistemic intuition' that Mary learns something. He goes so far as to propose that we may be right to be eliminativists about colour qualia, arguing that '…sensing red misrepresents how things are … If this is right, we should say that nothing is red, for nothing would be as our experience of red represents things as being;' (p. 432). The problem is that Jackson offers no solid account of the *way* 'our experience of red represents things as being', and this way (which Mary learns on first experiencing red) is at the core of the epistemic intuition in question.

Other contributions prove far more interesting and compelling. Essays by Lewis, Paul Churchland, Horgan and others focus on the epistemic issues, while Philip Pettit's contribution challenges the intelligibility of qualia, calling into question whether we are ever presented with color properties that can be distinguished from our capacities for and dispositions towards action and judgment. Daniel Stoljar examines the ontological commitments implied by acceptance of physicalism. He proposes a distinction between the dispositional properties identified by physical theory and the categorical properties belonging to paradigmatic physical objects, claiming that this allows us to reconcile our intuitions about qualia with acceptance of physicalism (of a certain type).

Taken together, the essays in this volume suggest that the greatest legacy of Jackson's argument is its motivation of much-needed critical reflection upon the meaning of physicalism and the commitments it actually entails. Robert Van Gulick's article is perhaps the most successful in bringing together the weighty epistemic and ontological issues raised by the history of the knowledge argument. Van Gulick exposes an often overlooked and deeply problematic assumption behind the knowledge argument — namely the assumption that all physical information is objective information. This assumption is behind the claim that it is possible in principle for Mary to have all of the physical information in her room, and Jackson's conclusion that the subjective information she learns upon release is not physical information. Van Gulick's proposal that some physical facts are irreducibly subjective brings the Mary controversy up-to-date, locating it squarely within the larger controversy between defenders of reductive and non-reductive physicalism. Will the intuitions behind the knowledge argument force us to stretch our notion of physicalism? Can our conception of the physical stretch far enough to accommodate them without becoming empty or indistinguishable from dualism? This volume is an excellent choice for anyone wanting to survey the recent history behind these important questions.

Rodrick Wallace
Consciousness:
A Mathematical Treatment of the Global Neuronal Workspace Model
Springer, 2005. 116 pp., $59.95
ISBN 0-387-25242-8 (hbk).

Alwyn Scott
University of Arizona

Among the most significant advances in twentieth-century science are Claude Shannon's information theory, which says that the amount of information carried by a message is proportional to the logarithm of its improbability and Kenneth Wilson's renormalization theory, showing that certain properties of some systems (crystals, fundamental fields, etc.) are independent of the scale at which they are described.

The stated aim of this little book is to use these ideas to provide a mathematical treatment of Bernard Baars' workspace model of consciousness that includes the hierarchical effects of embedding contexts — in particular, human culture. In the words of the blurb, the book:

> uses an extended statistical model of cognitive process, based on the Shannon-McMillan Theorem and its corollaries, to incorporate the effects of embedding physiological, social, and cultural contextual constraints which operate more slowly than the workspace itself, but severely limit the possible realms available to that workspace and hence to consciousness itself. The resulting 'biopsychosociocultural' treatment directly addresses criticisms of brain-only models of consciousness which have been raised in cultural psychology and philosophy, while remaining true to the current neuroscience perspective.

As the author, a research scientist at the New York Psychiatric Institute, points to Ruth Benedict's classic *Patterns of Culture* in support of the importance of cultural factors in consciousness and — properly, in my view — dismisses the role of quantum theory in mental phenomena, I approached the tasks of reading this book and writing a review with high expectations.

The book comprises six chapters plus a short appendix, with chapters one and six describing respectively the aims and results of the book. The second chapter reviews Shannon's information theory, oddly without making reference to his classic *Mathematical Theory of Communication*, with Warren Weaver. Motivated by the author's previous work in immunology, Chapter three discusses cognition as a generalized language in which the system is assumed to be Hamiltonian, so implying conservation of some energy-like quantity. Although such theories have been proposed in immunology, this is an unlikely property of open systems like the neocortex. Chapter four deals with means for describing dynamic thresholds, relating them to biological phase transitions; and Chapter five introduces some abstract geometrical concepts to formulate a very general representation of a hierarchical system, which can describe visual perception. In contrast to the complicated mathematical discussions of the inner three chapters, the appendix is short and simple but seemingly irrelevant.

In Chapter six, the author offers non-mathematical discussions of several ways in which the general hierarchical representation proposed earlier can be applied. These applications include to autocognitive development disorders (said to involve the immune system, tumors, fight-or-flight systems, regulation of blood pressure), plus emotions, sociocultural networks, obesity and schizophrenia. The examples are followed by an effective critique of those dynamic systems models of neural systems that ignore hierarchical structure.

Unfortunately, the book fails in two important ways. First, although the stated audience consists of those familiar with Baar's global workspace model and the 'current debate', it is not — in my view — appropriate for readers of the *Journal of Consciousness Studies*. Although I am familiar with both information theory and renormalization theory, and have long been a proponent of hierarchical structures in biological and cognitive systems, the author's arguments are difficult for me to follow. The three inner chapters, I suspect, will be impenetrable to the average reader of *JCS*. Second, editing and production of the book itself can only be described as sloppy. In support of the latter criticism, note the following.

(i) There is no index. (ii) While John Hopfield's relation between phase changes in physical systems and mental ideation is noted, and Hopfield is mentioned by name, no reference to his work is given. (iii) Paragraph indents appear below every formula, whether the following text picks up in the middle of a sentence, starts a new sentence, or opens a new paragraph. (iv) A vertical spacing of one and a half inches is typically provided for equations, but a forward slash is often used for fractions, making them more difficult to read. (v) The number for an equation at the bottom of page 54 appears at the top of page 55. (vi) Baars' name is misspelled in the second from last paragraph.

I could go on, but you get the idea — neither a copy editor nor a production editor got close to this project, which is a shame for two reasons. The author's ideas seem to be important for current developments in consciousness studies, and Springer is well known for its corporate ability to produce fine books.

BOOKS RECEIVED

Mention here neither implies nor precludes subsequent review.

Dorogovtsev, S.N. and Mendes, J.F.F., *Evolution of Networks: From biological nets to the internet and WWW* (Oxford University Press 2003)

Elsaesser-Valarino, Evelyn, *Talking with Angel: About Illness, Death and Survival* (Floris 2005)

Nunn, Chris, *De La Mettrie's Ghost: The Story of Decisions* (Macmillan 2005)

Suzuki, D.T., *Studies in the Lankavatara Sutra* (Kegan Paul International 1930/2000)

Walsh, Roger and Grob, Charles, S., ed. *Higher Wisdom: Eminent elders explore the continuing impact of psychedelics* (SUNY Press 2005)