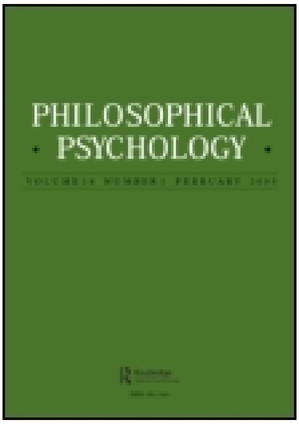

Philosophical Psychology
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/cphp20

An elusive target: A critical review of Clark Glymour's The mind's arrows

Brandon N. Towl, Jonathan Halvorson & Carl F. Craver

Available online: 19 Aug 2010

Routledge
Taylor & Francis Group

Book Reviews

Minds, causes, and mechanisms: a case against physicalism

Josep Corbí & Josep Prades
Oxford: Blackwell Publishers, 2000
ISBN 0631218025 (paperback, 256 pp., $33.95)

Minds, causes, and mechanisms, by Josep Corbí and Josep Prades, is a thoughtful critique of a widely held picture of causation and the way it gives rise to familiar puzzles about mental causation. Since the victory of materialist metaphysics in mid-century, philosophers of mind have struggled to find a picture of the mental that fits comfortably into a world understood as fundamentally physical. Giving a satisfactory explanation of the causal efficacy of mental phenomena like belief and desire is, and has always been, the most difficult part of this enterprise. No charge in philosophy of mind has carried greater sting than "epiphenomenalism," and yet it has been hard to see how an account of the mental which demands no more of the universe than is required to explain its non-mental aspects could vindicate the causal efficacy demanded by our conception of the mental. If there is no more to my desire, in some sense of "no more," than its physical basis, then how can my desire have any effects which are not already fully explained and accounted for by the effects of its purely physical underpinnings? The narrow path philosophers of mind must walk has been clear: give an account of the mental which squares its distinctive contribution to causal explanation with our physicalist presuppositions.

This stricture appears to permit only two strategies. Either one can claim that mental phenomena really just are physical phenomena, described in a different vocabulary, or one can try to make out a sense in which mental phenomena are distinct from physical phenomena but wholly dependent upon them. Donald Davidson's anomalous monism is the premiere example of the former strategy, while a variety of functionalist accounts vie for the title of "our most promising non-reductive account of the mental." *Prima facie*, both strategies appear to deliver the goods. Davidson's mental events just are physical events, governed by the same laws which govern the rest of the non-mental universe, while the explanatory power of functional states is wholly grounded by their realizing mechanisms in each particular case. In recent years, however, even such tailor-made solutions to the puzzle have been subject to criticisms that they do not, after all, provide a satisfactory causal role for the mental. On the one hand, the mental properties of Davidson's events seem causally irrelevant to their effects. On the other hand, functionalism has been subject to the twin charges (1) that the explanatory powers of functional characterizations are not only wholly grounded in but also wholly exhausted by that of their realizing mechanisms, and (2) that a functionalist account of contentful mental states requires a notion of narrow content to afford them genuine causal efficacy and that such a notion is unavailable. Whether or not these worries can be assuaged, it is clear that the problem of mental causation continues to dog the philosophy of mind as it has since Gassendi's critique of Descartes' interactionism.

While much recent literature is devoted to making good on these strategies, Corbí and Prades pursue a time-honored philosophical alternative: when a puzzle seems intractable, reexamine the assumptions which gave rise to it in the first place and argue that the puzzle is illusory. Motivated by the critiques of functionalism and monism, their target is what they call "causal physicalism," the familiar assumption that any genuine case of causation (mental, functional, etc.) must, on each occasion, be entirely dependent on the causal efficacy of some basic, physical properties. As proxies for this broad idea, Corbí and Prades identify three principles to which they believe the causal physicalist is committed: (1) intrinsicness, that two objects alike in their intrinsic properties share their causal powers in all contexts, (2) narrowness, that each causally efficacious non-intrinsic property must have a narrow correlate, and (3) minimality, that causally efficacious non-basic properties must supervene on a single, minimal collection of basic properties, all of which are necessary for the non-basic property's instantia-

ISSN 0951-5089/print/ISSN 1465-394X/online/03/010165–18 © 2003 Taylor & Francis Ltd
DOI: 10.1080/0951508032000067671

tion. They argue that all three principles are either false or can only be secured at unacceptable cost, such as ruling out the possibility of multiple realization for non-basic properties or requiring massive overdetermination of effects by causes. Their goal is to show the incoherence of causal physicalism and thus to expose it as a fraudulent standard for the causal efficaciousness of non-basic properties. Their criticisms also lead them to sketch, in the broadest terms, a more acceptable standard, one which it turns out mental properties have no trouble meeting.

The philosophical perspective Corbí and Prades bring to bear on these issues stems largely from a concern with the role of context in individuating the properties suitable for nomic generalization. They are among those who argue that the individuation of functional properties, and indeed of any properties which support nomic generalization and recognizably causal counterfactuals, proceeds ineliminably with reference to normal, actual conditions. In contrast, they maintain that causal physicalism requires an uncontextualized understanding of the way in which the causal powers of an object are a function of its intrinsic properties and the contexts in which it is placed, what they call a "metadisposition." They argue that such an understanding is unavailable in principle because any attempt to enumerate the properties of objects and contexts in a way which necessitates the effects fails, for there is always the possibility of additional contextual properties which would defeat the effect. The only way we, or even an omniscient being, could characterize genuine causal connections involves a perspective-bound approach and the way it eliminates deviant contexts from consideration. This argument forms the lynchpin of their critiques of standard approaches to understanding both strict and *ceteris paribus* laws.

In such a radically contextualist philosophical setting, the target principles struggle. Intrinsicness fails because intrinsic twins can have different causal powers in different contexts, even though one would have had the powers of the other in the same context. With intrinsicness goes strong supervenience, to be replaced by a weaker, "normal" supervenience claim, one restricted to spatio-temporal regions fit to provide the appropriate normal background conditions for individuation. Narrowness turns out to be incompatible with multiple realization because only exact physical duplicates share all their powers across all nomologically possible contexts, while the more generous individuatory practices offered by anchoring in normal and actual context vitiates their narrowness. Minimality runs afoul of strictures against overdetermination, as for any functional property, they claim one can pick out many compresent collections of properties, each of which suffices for instantiation of the functional property, and all of which compete to be the cause of the effect. With minimality goes the idea that every effect has a unique, complete physical explanation. In the broadest terms, Corbí and Prades aim to undermine the impression that the description of the world at the level of basic physical properties and strict nomic connections between them is a privileged one because describable from an absolute, uncontextualized perspective. The seemingly imperfect, perspectival understanding of functional, mental, and other non-basic properties and causal connections between them turns out to be the best understanding available for basic properties as well. Without this asymmetry, reasons for impugning the causal efficacy of the non-basic properties are said to vanish.

There are a number of points where readers may plausibly disagree with their reasoning or their formulation of what are supposed to be standard views. For example, their central argument against the availability of metadispositions appears not to take seriously the idea that principles which individuate functional properties should have the form of equivalence principles, a thought which should help insulate against the worry of contexts containing defeating properties. What they label "equivalence principles" are manifestly not such. So too is there a certain casualness with the notion of overdetermination. While they are explicitly aware that it is only independent causal chains which are meant to be ruled out, one worries whether their putative examples of overdetermination involve truly independent causal chains, especially when it comes to examples intended to undermine the minimality principle. Finally, while they often explicitly address the thought that these worries about context and individuation are merely epistemic, they are not fully successful in defusing them.

Minds, causes, and mechanisms is a difficult book. Much of this is due to the inherent complexity of the subject matter, but some is due to their prose. Although Corbí and Prades do a good job of taking the reader through the terrain of recent philosophy of causation and mind (absent, strangely, the influence of behaviorism) and regularly flag the structure of their arguments, the reader must often work hard to extract their reasoning, follow the flow of their thought, and discern the workings of their examples. The positive views to which their critique leads them are presented relatively briefly and it is hard to make out what they take the wider consequences of such an unorthodox view to be.

Corbí and Prades are engaged in a worthwhile project. The deepest presuppositions that frame debates on mental causation are too seldom dragged out into the light for questioning, and their critiques offer an opportunity to reassess much of what we often take for granted. They are aware that many of their claims will seem radical and work hard to give the reader a sense that there really are hard questions and puzzles where most see only obvious truths. They also perform a service in bringing to bear an underappreciated set of concerns—the interaction of context with individuation and causal explanation—which have not hitherto played a large role in the philosophy of mind literature. The problem of mental causation has remained intractable long enough for us to welcome any effort to reassess the ground rules which have persistently frustrated our hopes to find the mental a home in the physical world.

GUY ROHRBAUGH
Department of Philosophy
University of Utah
260 S. Central Campus Drive
Salt Lake City, UT 84112, USA
Email: rohrbaug@humnet.ucla.edu

Reclaiming cognition: the primacy of action, intention and emotion
RAFAEL NÚÑEZ & WALTER J. FREEMAN (Eds)
Bowling Green, OH: Imprint Academic, 2000
ISBN 0907845061 (paperback, xxix + 284 pp., $24.95)

When philosopher Rafael Núñez and neuroscientist Walter Freeman decided to name their recent collection *Reclaiming cognition*, they were quite clearly inviting a question: from whom does cognition need to be reclaimed? The answer to this implied question will be unsurprising to those who have followed recent debates in cognitive science. Cognition, Núñez and Freeman suggest, needs to be reclaimed from computational cognitive scientists, who claim thought is computation and who have been claiming for years that theirs is the only game in town.

The purpose of this collection to show that there are other games in town. The contributions are mostly from the outskirts of cognitive science, anti-computationalist radicals looking to pick a fight. Computational cognitive scientists raised on the milk of Chomsky and Fodor will find these essays frustrating for many reasons, among which are the following:

1. Computational cognitive scientists are mocked as "technocrats and analysts."
2. Computationalist research is not taken especially seriously, and is criticized from a feminist perspective.
3. Everything worthwhile is "embodied" or "situated" or "embedded," and computational cognitive science is not "embodied" or "situated" or "embedded."
4. Along with the usual anti-computationalist suspects (Gibson, dynamical systems theory, Heidegger and Husserl), Freud, Marx and Irigiray are favorably discussed.

Despite these rhetorical excesses, even an unfortunate use of the word "languaging," this collection has a lot going for it, and is deserving of a wide readership. Indeed, I will suggest that this is a book that the computational cognitive scientists especially really should read.

The supposed problem with computationalism, as anyone following current debates will have heard, is in its failure to address embodiment. The essay by Núñez describes three deepening levels of embodiment to which a theory of cognition might be committed. First there is *trivial embodiment*, which is simply the claim that thinking requires a brain. Everyone, even computationalists, believes this nowadays. The second level of commitment is to *material embodiment*. Commitment to material embodiment is commitment to trivial embodiments plus two more claims. First, cognition is decentralized. Second, cognition is constrained by the fact that it must be performed by an animal in real-time, with a brain like ours, in a body like ours, in a complex environment. In research on material embodiment, the focus is typically on low-level, not paradigmatically cognitive tasks such as motor control or feature detection. Finally, there is what Núñez calls *full embodiment*, which is the claim that *all* of our abilities are embodied in the second sense. That is, concepts, creativity, poetry, and the like also cannot be understood other than in terms of brains, bodies, and real-time actions in complex environments. Indeed, the thought is that this embodiment is the raw material for all of our "higher-level" abilities.

Thus, the problem with computational cognitive science from the point of view of the authors of the essays here is that computationalists are typically committed to just trivial embodiment, when cognition itself is fully embodied. Indeed, even the most computationalist-friendly thing here, the chapter by Andy Clark, adopts a commitment to full embodiment. Clark calls for a reconciliation of computationalism and fully embodied cognitive science. He suggests that we can take embodiment seriously, and still hold on the idea that cognition is computation by realizing that all mental representations, even those on which complicated computations are performed, are action-oriented. Clark is of course rarely considered a friend of the good old fashioned computational cognitive science, but in this collection, he seems stodgy, a reformer among revolutionaries. Most authors in the collection want no reconciliation, and argue that computationalism must be overthrown once and for all.

So, according to the contributors to this collection, cognition is to be reclaimed by practitioners of a non-computationalist, fully embodied cognitive science. But apart from the typically mocking rejection of computationalism and the commitment to full embodiment, is there anything like a unified story lurking in here? That is, is there one other game in town with sufficient explanatory power to be a competitor to computationalism? This collection is a mixed bag, but if you're selective, you can cobble together such a story. Doing so, however, requires ignoring several contributions, some of them very good, concerning topics such as logic (Longo), quantum mechanics (Gomatam), phenomenology (Goodwin, Sheets-Johnstone), and feminism (Rose). This is not intended to disparage these essays, which include some of my favorites of the collection, especially Hillary Rose's insightful, if ultimately self-contradictory, feminist critique of the current obsessions over consciousness.

What follows is an attempt at cobbling together such a unified story, which I will call "fully embodied cognition" (FEC). Start with the standard dynamicist view (see the essays collected in Port and van Gelder, 1995) that the brain, body, and environment are complex dynamical systems that are so tightly coupled as to be inexplicable as separate entities. Understanding the mind, the task of cognitive science or any scientific psychology, just is the understanding the interactions of these dynamic systems (Cisek, Iverson & Thelen, Turvey & Shaw, Shaw & Turvey; Hardcastle, echoing Griffiths, 1997, suggests that emotions should also be understood dynamically, at least developmentally). The activity of the brain is also to be understood as a dynamical system, to be explained using the mathematical tools of dynamical systems theory (Freeman, Nicolis & Tsuda, Cisek). Its main task is to act as the controller for the body in the environment (Cisek).

This coupling (brain-body-environment) determines the content of many or most of our perceptions and thoughts. The most basic cases of perception are action-oriented, and the percepts are affordances (Gibson, 1979) or opportunities for activity. Furthermore, much of our action is what Edward Reed (1996) called exploratory action, action designed to facilitate more felicitous perception (Cisek, Clark, Turvey & Shaw, Shaw & Turvey). There is, according to this view, a very tight coupling between perception and action: (some? most?) perception is action-oriented and (some? most?) action is perception-oriented.

Of course, not all of human thought is action-oriented. To explain paradigmatically cognitive activities like planning and doing math, activities in which we are in one way or another decoupled from the environment, FEC invokes concepts. First, because we are dealing with a story that focuses on the primacy of embodied action, concepts will be action- and body-based (Núñez) in a way that is familiar to readers of Lakoff and Johnson (1980, 1999). More importantly and less familiarly, the job of concepts, according to this picture, is to create and enforce an artificial separation of mind and world, to hold the world at arm's length briefly, imperfectly, in order to think about it (Skarda, Rosch).

FEC is, I think, a very compelling story. Perhaps it is even sufficiently powerful to compete with computationalism. There are two ways in which this story is striking. The first is that FEC reflects the ongoing renewal of interest in Gibsonian ecological psychology (Gibson, 1979) in cognitive science circles. (Note however that ecological psychology has been around and thriving for quite some time; it's just that cognitive scientists are coming back around to it.) The second thing that's striking about it is that FEC is very nearly identical to Brian Cantwell Smith's theory of registration (1996), a theory of representation cum ontology designed to account for intentionality in both humans and machines.

The similarity between FEC and Smith's theory of registration points to the reason that this collection ought to be of interest to the computationalists from whom it aims to reclaim cognition. Smith, a lapsed computationalist, suggests that the theory of registration follows from careful years of research on computation and attempts to carefully spell out the computationalist hypothesis that the mind is a computer. That is, if one tries really hard to do non-embodied cognitive science, one ends up concluding that the account of cognition one can cobble together from *Reclaiming cognition* is correct. Perhaps, then, computationalists will end up here after all. They could save time by reading this collection sympathetically, along with Gibson (1979) and Smith (1996).

Acknowledgement

Thanks to Roger Thompson for comments on an earlier draft of this review.

References

GIBSON, J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

LAKOFF, G. & JOHNSON, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

LAKOFF, G. & JOHNSON, M. (1999). *Philosophy in the flesh*. New York: Basic Books.

PORT, R. & VAN GELDER, T. (1995). *Mind as motion*. Cambridge: MIT Press.

REED, E.S. (1996). *Encountering the world*. New York: Oxford University Press.

SMITH, B.C. (1996). *On the origin of objects*. Cambridge: MIT Press.

ANTHONY CHEMERO
Scientific and Philosophical Studies of Mind Program
Franklin and Marshall College
Lancaster, PA 17604–3003, USA
Email: tony.chemero@fandm.edu

Brave new mind: a thoughtful inquiry into the nature and meaning of mental life

PETER DODWELL

Oxford: Oxford University Press, 2000
ISBN 0195089057 (hardback, xiii + 250 pp., $35.00)

In this quirky, contradictory book, Dodwell sets out to "mount an even more radical challenge to the standard model [of explanation in cognitive science] than Penrose" (p. 81). But he also says that his "goal is not to attempt to demolish the standard model, but rather to raise the level of awareness of the consequences of holding it" (p. 106). Dodwell talks about the clockwork nature of the standard model, and says that it cannot explain cognition (which he calls "mentation"), but he also says that he "does not seek to undermine the achievements of cognitive science, only to put them in proper perspective" (p. 195). He says that cognitive scientists and practitioners of artificial intelligence (AI) focus on "the world of material objects and processes, where physical law holds sway" (p. 155), but then says that AI practitioners focus on the software, which they liken to *mind*. He argues for the independence of levels of explanation and endorses what he calls the "Peters principle"—that "no statement of psychological import can be deduced from premises that themselves have no psychological reference" (p. 191). Thus, the mental cannot be explained in terms of the physical. Initially, Dodwell says that it can only be explained in terms of the mental, but later says that it should be explained in terms of objective knowledge and ideals. This, however, violates the Peters principle.

The structure of the book is as follows. Chapter 1 says that philosophers of cognitive science have a widely accepted, although seldom explicitly stated, a metaphysical position—"the metaphysics of materialism, plain and simple" (p. 8). This is an outdated, "clockwork" view of the world that is essentially the science of the 1860s. Chapters 2–3 discuss the disciplines that have fed into the "standard model" in cognitive science: experimental psychology, neuroscience, AI, linguistics, and philosophy of mind. Chapter 4 lists the "axioms" of this model, which "collectively define the metaphysics of cognitive science" (p. 58). These include:

(1) mental activity is a function of brain processes,

(2) the physiology of the brain will eventually be understood in terms of chemical and physical laws, so that, ultimately, statements about mental activity will be reducible without remainder to statements couched in physical terms, and
(3) the brain is a product of the natural biological world, developed under the same evolutionary pressures as the rest of the body.

The "recent computational trend in cognitive science" (p. 59) has added some widely accepted beliefs to these axioms, including:

(4) the goal of cognitive science is to "determine what sort of machine the brain is" (p. 59), and
(5) theories must be explicit enough to be "expressed in terms of *efficient procedures* for implementation, i.e., as algorithms, either in abstract, formal terms, or as simulations on a machine" (p. 59).

Dodwell argues against this materialism using Popper's distinction between worlds 1, 2, and 3. *w1* is the world of physical objects and events. *w2* is the world of mental contents—ideas, inner thoughts, etc. *w3* is similar to Plato's World of Forms, except that most of its contents are the products of human contrivance. Following Popper, Dodwell argues that *w3* causally influences *w1*. When we discover a proof, the proof was there *waiting to be discovered*, and, once discovered, it causally influences *w1*, in terms of what goes on in brains and on paper. Moreover, it does so via *w2*, by changing people's beliefs, etc. Consequently, *w2* causally effects *w1*, so that it is not just an epiphenomenon of *w1*. Exit materialism.

Chapter 5 argues for the independence of levels of explanation. Dodwell distinguishes between three levels of perceptual activity: sensory physiology, perceptual organization and awareness, and cognitive use. "Each level has its own appropriate theoretical constructs, specific models and mechanisms of operation" (p. 98). He then distinguishes between prediction and causation on the one hand and explanation on the other. Temperature *t* and pressure *p* will cause a snowflake of type *s* to form on a nylon thread suspended in a super-cooled vessel of water vapor. *t* and *p cause s*, and we can predict that *s* will occur given *t* and *p*; but *t* and *p* do not *explain s*. For this we would need a model of molecular dynamics from which the characteristics of the snowflake could be deduced. Explanations are level-specific. Information about one level might enable us to predict what will happen at a higher level, but cannot explain it. Dodwell predictably applies this distinction to the mind–body problem: sound waves striking the tympanum cause us to hear a sound, but do not explain why we hear it, because we do not have a theoretical mechanism that enables us to deduce what is heard from what is going on in the ear and brain (p. 95).

From this point onwards, I find the argument hard to follow. Chapter 6 argues against "monolithic theories of perception" (p. 108), of the sort espoused by Gibson. Chapter 7 is a detour into mathematics, aimed at showing that we cannot understand "mathematics and how it is done" (p. 150) by studying the brain. Chapter 8 says that much is wrong with the standard model, including a failure to distinguish between prediction and explanation, and needing the categories of folk psychology whilst believing they must be discarded as pre-scientific.

Chapter 9, unfortunately called "The sacred river," sets out Dodwell's alternative framework. As far as I understand it, "the sacred river" is *w3*, now construed in a way more reminiscent of Hegel than Popper. Popper rejected Hegel's notion that *w3*, which Hegel called "Spirit," evolves under its own steam, and was at pains to stress that most of the contents of *w3* are the products of human effort. Dodwell turns this about: "rather than *w3* being just a world of intellectual and other cultural products, it is seen to be the progenitor? of all that is culturally powerful and precious in humankind. We thus come to see ourselves as the vehicles for the expression of the entities of *w3*, not—or not only—as their creators. [The human mind is] the bearer of … ideas and ideals that are gestated in *w3* but are brought to birth out of the flashing stream" (p. 192). All genuine psychological explanation "thus seems to require us to enter, to draw our ideas from, *w3*" (p. 192). Surprisingly, Dodwell does not mention Hegel.

What are we to make of all this? First, Dodwell's axiomatic characterization of cognitive science is quite strange. It is not clear where the axioms come from. He has surveyed the disciplines that feed into cognitive science, but has not derived the axioms from the survey. Nor is it clear that the axioms are correct. What most obviously characterizes cognitive science is its use of computational models of cognition that enable us to test theories about what would otherwise be the "black box" of the mind (sometimes called "cognitivism"). Dodwell cites it, not as an axiom, but as a "widely accepted belief" that came about with the "recent computational trend in cognitive science" (p. 59). I do not think that there was any cognitive science before cognitivism. Cognitive science came about when different disciplines realized that they were using computers (or at least formal, algorithmic methods, as Dodwell says) to model cognition. So is cognitivism committed to materialism? Surely not. Cognitivism and its close cousin, classical, symbol-handling AI, are concerned with the software, not the hardware. The software runs on silicon computers in the same way that the mind "runs" on the carbon wetware of the brain. It is software, or mind, that matters, and the physical nature of the hardware is irrelevant, so long as the software can run on it. Dodwell comes close to saying this when he says "theories must be expressed in terms of efficient procedures for implementation, i.e., as algorithms," (p. 59), and he openly admits it later on: "Computer scientists who simulate mental functions generally claim that it is the program they devise that is crucial? The hardware [is] of no theoretical importance in the sense that many different types of system could be made to run the same simulation, including, for one of this persuasion, the 'wetware' of the brain" (p. 158). All that he has to say about this is that it "should give one pause for thought" (p. 158). Ironically, he seems to have got his levels of description wrong.

There are other ways in which Dodson misrepresents AI and cognitive science. He says that cognitive science has "virtually nothing" to say about creativity and imagination (p. 137), and that it treats the "revelations" of the creative process with "a shrug of the shoulders" (p. 150). This is simply not true, for the study of creativity is a subfield in cognitive science, with its own literature and conferences. A good introduction is Boden (1990), which amongst other things outlines the programs that have been written to model theory formation in mathematics and the sciences. A great deal of attention has been given to trying to understand the apparently "revelatory" nature of the creative process.

Let us return to Dodwell's main attack on cognitive science. He says that psychological explanations cannot be deduced from non-psychological premises, so that we cannot explain cognition in materialistic terms. In that case, stubbing my toe against a rock does not explain the pain in my toe, and drinking all that whisky does not explain why the world is spinning around, or why I will feel dreadful in the morning. There may be some sort of theoretical or explanatory gap between the physical stimulus and the ensuing experience. What worries me is that this causes Dodwell to shy away from trying to explain any of our experiences in terms of our interactions with the world. Instead, we should limit ourselves to explaining cognition in cognitive terms (though this is not what he actually does, as we shall see). But how do we explain the pain *without* referring to the rock, or the dizziness *without* referring to the alcohol? First, I do not see how this can be done. Second, the *natural* explanation is in terms of the rock and the alcohol. This is how our folk psychology works, and Dodwell stresses the importance of folk psychology, which, he says, names and describes the cognitive phenomena we are trying to explain. Although he is equivocal about the relationship between folk psychology and the standard model, he says that if the standard model clashes with folk psychology, so much the worse for the standard model. Equally, then, if his account of explanation clashes with folk psychology, so much the worse for his account of explanation.

Dodwell writes that he wishes to create "a rich and full picture of humankind" (p. 30). I think that a rich and full picture of humankind locates us in the physical world (one is inclined to say, "the natural world") that is our natural home. It is not unreasonable to say that AI and cognitive science have had a humanizing effect by showing that attempts to model intelligence in terms of disembodied symbol-handling do not succeed, and that we need to be situated in the world to acquire the low-level skills that underlie and drive cognition. Ironically, symbolic AI started out with the kind of profoundly non-biological approach to intelligence that Dodwell seems to be advocating, and to some extent discovered the importance of embodiment. Dodwell's explanation would take us back to the disembodied approach.

When the chips are down, however, Dodwell does not talk about explaining cognition in cognitive, *w2*, terms. He talks about explaining it in *w3* terms. Why doesn't he think that this violates the Peters principle? First, he moves from folk psychology to what he calls its "grand embodiment— the flower of civilisation … including the great cultural achievements of the past three thousand years" (p. 197). Second, he apparently thinks that *w3* entities are psychological because they are intentional. I think he is wrong about this. Something can be intentional, in the sense of "being about something," without being psychological or cognitive, in the sense of "being a cognitive state" or "being in a mind." Consider the belief "beer is good for you." In one sense of "belief" this is a psychological state that I might have and you might not. Beliefs in this sense are the furniture of our mental lives, and our different beliefs will to some extent determine the psychological differences between us.

But in another sense, "beer is good for you" is not a psychological state. It is an *objective proposition* (we sometimes say it is the *content* of belief). *w3* beliefs are beliefs in this sense. They exist independently of us, and we believe them or do not believe them, just as we see or do not see objects in the physical world. In this sense, beliefs are denizens of another world, and invoking them to explain *w2* violates the Peters principle just as much as invoking physical objects. But invoking physical objects at least enables us to tell a causal story about why we hear and see things. We cannot tell such a story about the contents of *w3*. They do not cause us to believe them.

Dodwell talks about the importance of idealization and admires Chomsky's concept of the "competence of an Ideal Speaker," who knows its language perfectly (he also admires MacNamara, but does not refer to MacNamara's concept of an "Ideal Reasoner"). I think we need to be careful about the role of idealization in here. Saying that we should study language by studying the competence of the Ideal Speaker gives us the impression that we are studying mind and doing psychology, when we are really only studying language. It is like saying that we should study the universe as God sees it—which is just another way of saying that we should study the universe as it really is.

I do not like to write negative reviews, but I find it hard to be charitable in this case. Dodwell takes forever to state his case, and it is never entirely clear what the case is. The Peters principle, a vital plank in his argument, is not introduced until late (p. 166), and not fully stated until even later (p. 191)—nine pages from the end of the final chapter. As well as the inconsistencies cited in the first paragraph, he also says "the standard model does actually accept … folk psychology" (p. 154), but it believes that folk psychology "must be disregarded as 'pre-scientific'" (p. 166); the standard model can be reconciled with folk psychology, but this "requires some thought," which he does not provide, because this is the end of the book. He calls for a "greater humility" and "proper modesty" in explaining mind (p. 196), but he tends to be arrogant. In the early chapters he repeatedly appeals to a work of his own that is in preparation. He says that his distinction between a cause and an explanation goes against "a strong tradition in the philosophy of science" (p. 91), but says, "… too bad! It is an unfortunate tradition!" (p. 91).

There are typographical errors on the following pages and lines: viii/20; 5/20; 9/27; 12/32; 20/4, 6; 21/27; 23/33; 27/36; 69/29; 168/30; 200/14; 203/44; 204/42; 207/1, 3; 209/33 (twice); 214/12; 218/13. There is no reference for the quotation on page 61. "Hofstadter" is spelled "Hofstaedter" throughout.

References

BODEN, M. (1990). *The creative mind: myths and mechanisms*. New York: Basic Books.

TERRY DARTNALL
School of Computing & Information Technology
Griffith University
Nathan, Brisbane
Queensland 4111, Australia
Email: terryd@cit.gu.edu.au

Dynamical cognitive science

LAWRENCE M. WARD
Cambridge, MA: MIT Press, 2002
ISBN 0262232170 (hardback, 371 pp., $45.00)

Ward's landmark book *Dynamical cognitive science* provides a comprehensive introduction to the application of dynamical systems theory in the field of cognitive science, as well as an outstanding synthesis of interdisciplinary theories and methods. Concepts and techniques of dynamical systems science are made available to cognitive scientists, behavioral scientists and neuroscientists. The book also presents important epistemological views emphasizing the role of time and change—the "unfolding in time"—in several scientific domains.

Given the current scenario in cognitive science, Ward's book explicitly and implicitly highlights an important alternative approach to currently unsolved problems in the field. One of these core problems is the emergence of macroscopic behavioral and cognitive states from microscopic neural interactions, and from spatio-temporally local input or internal states. An appropriate characterization of macroscopic states in neuro-cognitive systems in terms of dynamic systems can also contribute to bridge the foundational gap between subsymbolic and symbolic representational domains. Another crucial problem in this framework, which can be successfully approached in terms of dynamic systems theory, is modeling the reciprocal interaction between neuro-cognitive systems and their environment. In general, a dynamic system approach is ideal to theorize and model autonomous processing and representational dynamics "going beyond the stimulus given."

In *Dynamical cognitive science*, Ward successfully makes relatively technical concepts or analytic tools understandable even to non-specialist readers, without any loss of scientific rigor and formal elegance. Despite the apparent heterogeneity of the relatively short yet dense 35 chapters, their logical fluency and thematic integration are remarkable. The unavoidable space limits are generally counteracted by an accurate selection of the most relevant contents about the topic of each chapter. The reader is appropriately referred to other "dedicated" sources throughout the book.

Elegantly, Ward's book starts with introducing the fundamental dynamic concepts of order, rhythm and change, which are transposed from the primordial domains of magic and rituals to formal dynamics. The relevance of a dynamical approach to behavior is convincingly introduced with reference to psychophysical laws, and specifically in discussing the limitations of Steven's law (Steven, 1975), neglecting both the amount and the timing of fluctuations over trials, as compared to Fechner's approach (Fechner, 1860), using the *amount* of fluctuations to measure sensation, and finally to a more comprehensive dynamic approach also taking into account the *timing* of fluctuations in stochastic and deterministic processes (e.g. Kelso, 1995).

In Chapter 2, the serial nature of the universe and human behavior (referring to Lashley's problem of serial order in behavior), as well as the Markovian analysis of behavior in terms of probabilistic transitions depending only on the current state, is introduced. In Chapter 3, the rhythmic nature of interpersonal behavior, music and brain activity is discussed. Haken *et al.*'s (1985) model of coordination is presented as an example of *synergetics*, the interdisciplinary theory of pattern formation developed by Haken (1983). Different concepts of time, space–time and time measure are considered in Chapter 4, ending with an emphasis on *discretized time* in reformulating some domains of physics and possibly quantum gravity. These three chapters exemplify the interdisciplinary character of Ward's book.

In one of the key chapters, Chapter 5, and following the general conceptual analysis of time in the previous chapter, time is considered with reference to cognitive processes, in terms of "temporal unfolding of cognitive behavior" (p. 35). After having introduced the concept of finite state machine (giving rise to sequences with a different complexity), in this chapter Ward presents a variety of examples of time-variant psychological states, like emotional states evolving over different time-scales, visual adaptation, scanning by eye movements, learning and problem solving. The temporal relationships between stimuli in visual masking and attention are also mentioned. Finally, the constructive and dynamic (in terms of network models) properties of memory are shortly discussed. Perhaps, given the complexity of each of these cognitive domains, more than one chapter could have been dedicated to their temporal properties.

Ward's book may also be insightful if it is read in an epistemological perspective. For instance, in Chapters 6–10 the book introduces general approaches and concepts, with several epistemological implications, like General Systems Theory, Dynamical Systems Theory, the definition of formal theory and the principle of complementarity. Given the aims of Ward's book, Chapter 8 appears particularly important, as it puts forth the main differences between *dynamical* models, concerned with a succession of states, and *statical* models, which implicitly or explicitly assume that the relevant state of a system is constant. Appropriately, Ward claims that "most theories and models in psychology are statical models" (p. 62). Structural models emphasizing the *connectedness of elements*, with special reference to *graph theory*, as well as their relationships to dynamic models, are discussed in Chapter 9. At a crucial point of the book, Chapter 10 clearly defines deterministic and stochastic models, concluding that these "two types of models can lead to different outcomes even when applied to the same data," and that "useful models sometimes have both deterministic and stochastic parts" (p. 87).

Given these premises and definitions, the subsequent eleven chapters are concerned with properties and models of stochastic processes. Although some of these chapters do not directly refer to cognitive or neural processes, they describe concepts and analytic techniques that are fundamental in understanding other more cognitively- or neurophysiologically-oriented chapters of the book. One of these chapters (Chapter 13) is concerned with stochastic models in physics, with the crucial distinction between different description levels (microscopic, mesoscopic and macroscopic).

Chapter 16 is focused on 1/*f* or "pink" noise, a temporally structured noise supposed to represent the signature of a complex system (Waldrop, 1992) in human cognition. Several studies on the role of 1/*f* noise in perception and behavior (e.g. estimation of time intervals, reaction times) are presented. Complementarily, Chapter 17 is on 1/*f* noise in the brain, in magnetoencephalographic, electroencephalographic and event-related potential recordings. Provocatively, on the basis of several experimental findings, this chapter concludes that brain activity is characterized by 1/*f* noise at different levels, from neuronal currents to macroscopic activity evoked by external stimuli. However, as a reply to Ward's claim, it may be pointed out that much more neurophysiological and neurocomputational evidence is necessary to specify and constrain the neural mechanisms behind these potentially crucial 1/*f* dynamics.

Given the evidence of 1/*f* noise in cognitive/behavioral and neural processes, Ward considers some models and the statistical theory of 1/*f* noise in Chapters 18 and 19, respectively. Especially interesting is *self-organized criticality* (Bak, 1990), observed when a dissipative system with many degrees of freedom operates near a point of instability (criticality) in non-equilibrium conditions, without driving influences from the external (and therefore by self-organization). The potential implications of self-organized criticality for neural computation with spiking neurons were also put forth by Hopfield (1984), one of the pioneers of neural network theory. Along with Haken's synergetics mentioned in Chapter 3, self-organized criticality may be considered as one of the most versatile interdisciplinary theories in the dynamic systems framework, with enormous implications for the explanation of neural and cognitive phenomena on different spatial and temporal scales.

Another core concept in dynamic systems theory, with important implications for neural and cognitive modeling, is *stochastic resonance*, considered in Chapters 20–22. As explained by Ward, stochastic resonance is a nonlinear cooperative effect characterized by an amplification of weak random signal fluctuations by an independent weak (often periodic) signal. The relevance of stochastic resonance is discussed in Chapter 21, whereas Chapter 22 discusses the characterization of single-neuron and neuronal network activity in terms of stochastic resonance.

Finally, in Chapter 23 the concept of chaos is introduced. In Chapter 24 the distinction between chaos and randomness is clarified, and the logistic difference equation introduced. Recently, in an attempt to develop a neurocomputational architecture with multi-domain implications, Coupled Map Lattices (Kaneko, 1990) governed by logistic equations have been suggested to model several perceptual and memory processes, in terms of chaotic patterns of *spatio-temporal coherence* (Raffone & van Leeuwen, 2001; van Leeuwen & Raffone, 2001; van Leeuwen *et al.*, 1997). Given the computational efficiency in their implementation, these models also appear promising in real world applications, as in the segmentation of medical images.

Chapter 25 clearly introduces and characterizes nonlinear time series analysis, which can be usefully applied in neuroscience and behavior analysis (see Heath, 2001, for an encompassing dedicated book). Chapters 26–30 interestingly relate the concepts and analytic tools presented in the previous chapters to behavior and brain studies. In Chapter 27, for instance, the evidence of chaos in the brain is discussed (see also Tsuda, in press, for a recent theoretical synthesis about chaotic patterns in brain function).

In Chapter 31, one of the most important of the book for neural modeling implications, Ward claims that *relaxation oscillators*, characterized by a quick jump followed by a slow change over a so-called "relaxation time," may represent a foundation for dynamical modeling. This view appears consistent with the experimental evidence of cognition-relevant oscillations and coherence in neural systems. Neural relaxation oscillations and synchronization dynamics may play a crucial role in binding local neural activities in the brain into larger representational units (see Singer, 1999, for a review).

The last four chapters present synthetic overviews on the evolution and ecology of cognition, dynamical cognitive neuroscience, dynamical computation and dynamical consciousness. These final chapters of the book clearly put forth the broad implications of dynamical systems theory and methods for different core domains of neural and cognitive modeling.

However, it may be observed that Ward's book (mostly in Chapter 34) does not sufficiently consider the *connectionist approach* to cognition (e.g. Rumelhart & McClelland, 1986), which provided a first important dynamic approach to neuro-cognitive information processing. Generally, it appears somewhat unclear to what extent the current models based on dynamical systems theory can explain and mechanistically simulate the learning processes crucial to behavior and cognition, as well as characterizing their neural correlates. One of the exceptions in this scenario may be represented by Tsuda's (in press) brain theory based on *chaotic itinerancy*.

A dynamical system approach, and more specifically models based on networks of coupled neural oscillators, is one of the best candidates to solve the following high-level representational problem in the brain: how do complex mental processes and conscious awareness emerge from microscopic interactions between neurons in the brain? Several neuroanatomic and neurophysiological studies demonstrated that neurons in the cerebral cortex have access to limited fragments of information from the external world. For instance, neurons in the visual cortex are specialized in processing some attributes or features of visual stimuli, like specific orientations, colors or shapes. However, despite this functional segregation in the visual cortex, our perceptual experiences are integrated and coherent, without any awareness of this microscopic diversity and separate representations in our brain.

The processing specialization in the cortex leads to the crucial problem of integrating the outputs (signals) from separate neural detectors or processors, each of which encodes just a fragment of the whole picture. In neuroscience and cognitive science, this problem is commonly addressed as the *binding problem*. The binding problem can be generalized and thus include the integration of signals from different brain regions. The following passage from Adina Roskies (1999, p. 8) eloquently points out its broad implications:

> How does something as simple and mechanistic as neural firing add up to subjectivity, raw feelings, a self? Are the mechanisms that allow us to attribute the correct color and shape to an object the same ones that lead to the unity of phenomenal experience? Will the solution of the binding problem be the solution to the mystery of consciousness? (Roskies, 1999, p. 8)

As discussed in Ward's book, it has been suggested that the binding problem can be solved by temporally-correlated firing of neurons coding for different properties of a given stimulus (Singer, 1999; von der Malsburg, 1981), even during visual short-term storage after stimulus offset (Raffone & Wolters, 2001). Chaotic neurodynamics may be necessary to the self-organization of binding codes at different processing stages. For instance, in simulation studies based on core dynamic systems principles, van Leeuwen and Raffone (2001; see also Raffone & van Leeuwen, 2001) have recently demonstrated that chaotic neural synchronization patterns are dynamically flexible to represent efficiently complex visual scenes or active memories in retrieval, even when network units code for more than one active pattern. However, many foundational and technical questions about neural representations and cognitive modeling remain to be answered, and probably only the combination and integration of dynamic system and connectionist methods will provide the necessary answers in the future.

References

BAK, P. (1990). Self-organized criticality. *Physica A*, *163*, 403–409.

FECHNER, G.T. (1960). *Elemente der Psychophysik*. Leipzig: Breitkopf and Härtel.

HAKEN, H. (1983). *Synergetics*. New York: Springer.

HAKEN, H., KELSO, J.A.S. & BUNZ, H. (1985). A theoretical model of phase transitions in human hand movements. *Biological Cybernetics*, *51*, 347–356.

HEATH, R.A. (2000). *Nonlinear dynamics: techniques and applications in psychology*. Mahwah, NJ: Lawrence Erlbaum Associates.

HOPFIELD, J.J. (1984). Neurons, dynamics and computation. *Physics Today*, *47*, 40–46.

KANEKO, K. (1990). Clustering, coding, switching, hierarchical ordering and control in a network of chaotic elements. *Physica D*, *41*, 137–172.

KELSO, J.A.S. (1995). *Dynamic patterns: the self-organization of brain and behaviour*. Cambridge, MA: MIT Press.

RAFFONE, A. & VAN LEEUWEN, C. (2001). Activation and coherence in memory processes: revisiting the Parallel Distributed Processing approach to memory retrieval. *Connection Science*, *13*, 349–382.

RAFFONE, A. & WOLTERS, G. (2001). A cortical mechanism for binding in visual working memory. *Journal of Cognitive Neuroscience*, *13*, 766–785.

ROSKIES, A.L. (1999). The binding problem. *Neuron*, *24*, 7–9.

RUMELHART, D.E., MCCLELLAND, J.L. (1986). *Parallel distributed processing: explorations in the microstructure of cognition*, Vol. 1. Cambridge MA: MIT Press/Bradford Books.

SINGER, W. (1999). Neuronal synchrony: a versatile code for the definition of relationships? *Neuron*, *24*, 49–65.

STEVENS, S.S. (1975). *Psychophysics: Introduction to its perceptual, neural, and social prospects*. New York: Wiley–Interscience.

TSUDA, I. (in press). Towards an interpretation of dynamic neural activity in terms of chaotic dynamical systems. *Behavioral and Brain Sciences*, *24*.

VAN LEEUWEN, C. & RAFFONE, A. (2001). Coupled nonlinear maps as models of perceptual pattern and memory trace dynamics. *Cognitive Processing*, *2*, 67–111.

VAN LEEUWEN, C., STEYVERS, M. & NOOTER, M. (1997). Stability and intermittency in large-scale coupled oscillator models for perceptual segmentation. *Journal of Mathematical Psychology*, *41*, 319–344.

VON DER MALSBURG, C. (1981). The correlation theory of brain function. Internal report 81–2. Göttingen: Max Planck Institute for Biophysical Chemistry.

WALDROP, M.M. (1992). *Complexity: the emerging science at the edge of order and chaos*. New York: Simon and Schuster.

ANTONINO RAFFONE
Department of Psychology
University of Sunderland
St. Peter's Campus
Sunderland SR6 0DD, UK
Email: antonino.raffone@sunderland.ac.uk

Consciousness and cognition

MICHAEL THAU
Oxford: Oxford University Press, 2002
ISBN 0195141814 (hardback, 304 pp., $39.95)

Central to the debates over consciousness these days is the question of whether perceptual awareness can be accounted for solely in terms of the representational content of perceptual states. Representationalists like Michael Thau answer this question in the affirmative. The view of the so-called "qualia freaks" is that what it's like to have experiences outruns their representational content. Qualia freaks like Block take inverted spectrum scenarios to be evidence for their views. Consequently, many representationalists such as Dretske, Harman, Lycan, and Tye spend time trying to show that these cases are impossible as described by the qualia freak. *Consciousness and cognition* occupies a middle ground in that it makes room for spectrum inversion within a representationalist account of perceptual awareness. In that regard, Thau's work is indebted to Sydney Shoemaker's, in spirit if not in its details.

In inverted spectrum cases, two subjects are supposed to be alike with respect to the colors that they perceive objects to have, but systematically differ with respect to what it is like for them to perceive objects as having colors. Subject S perceptually represents objects to be red and green, and what it is like for S to do so is R and G, while what it is like for S^* to perceive red and green, respectively, is G and R, an inverted version of what it is like for S. Neither subject misrepresents the colors of objects, and since those colors are all that is represented in such perceptual episodes the differences between S and S^* with respect to R and G must be non-representational. Therefore, the argument goes, there are aspects of conscious awareness that cannot be explicated in terms of the representational content of perceptual states and representationalism is thus false.

One of Shoemaker's ideas is that the representation of redness, squareness, and the like does not contribute to the phenomenal character or the "what it's like" of experiences. Other perceptually represented properties that are correlated with redness, etc. determine what it is like to have those experiences. Thau agrees, but unlike Shoemaker, he claims that the paradigmatically perceptible properties like redness, squareness, distance, brightness, loudness, pitch, smells and the like are not perceptually represented at all, and he uses a variant of Jackson's knowledge argument to establish this negative claim. Briefly reviewing the argument is a useful way of getting a feel for the book as a whole.

Imagine Mary in her black-and-white room being shown a red object for the first time without being told that the object is red. Upon seeing it she wonders, "Do fire engines have that property?" She is then told that the object is red, and exclaims, "So fire engines have that property!" despite having known all along, we can assume, that fire engines are red. Thau believes that the property Mary ostends—*that property*—cannot be redness and thus that Mary does not perceptually represent the object as being red. The following seems to be an accurate reconstruction of his argument:

(1) Mary gets new information when she is told that the object she sees is red.

The evidence for this claim is the fact that she can wonder, before being told that the object is red, whether fire engines have *that property*. Being told that the object is red answers her question, so it must constitute some new information.

(2) Whenever someone gets new information in virtue of accepting some sentence, she also gains a belief that differs from any of her old beliefs with respect to its *what*.

This is what Thau calls the New Information Principle, which is central to his defense of a pure Millian approach to mental representation in Chapter 3. Beliefs differ with respect to their *what* just in case they differ in what they are about, as opposed to differing in how they are about it. For Thau, this is the only respect in which beliefs qua beliefs can differ from one another. Thau incidentally thinks that even if his arguments in favor of pure Millianism are not decisive, the New Information Principle stands and allows this argument about Mary to go through (pp. 185–187). If Thau is right of course, then his book hangs together much more loosely than one would have hoped. We are given an argument for pure Millianism and a relatively independent argument about the nature of perceptual states *vis-à-vis* the beliefs to which they give rise. Leaving that point aside, the argument continues:

(3) Thus, Mary's exclamation registers a new belief.

Registering a new belief is not tantamount to expressing a new belief, and Chapter 4 is a nice discussion of the semantic value of sentences and the information that they are used to convey *vis-à-vis* Grice's work on implicature. One can register a new belief without expressing it directly, and the issue is whether her exclamation registers such a belief, and if so, what the belief amounts to.

(4) Thus, this new belief differs from her belief that fire engines are red with respect to its what.
(5) If redness = *that property*, then the difference between her new belief and the belief that fire engines are red must be wholly explicable in terms of some descriptive mode of presentation that Mary did not previously associate with the term "red."

In order for the newly registered belief to differ in its what from the belief that fire engines are red, it must do so by associating with "that property" some description that picks it out that is distinct from any other description Mary previously associated with "red." Otherwise, one could find a belief of hers that is identical with respect to its *what* and thus identical *tout court* to the belief that she has registered. Remember that on Thau's account if *that property* = redness, then the belief expressed by Mary—"Fire engines have *that property*"—just is the belief that fire engines are red, though the belief *registered* by the former may be distinct from that registered by the latter. Moving on, Thau thinks that the consequent of the foregoing conditional is false:

(6) The difference between the beliefs is *not* wholly explicable in terms of some descriptive mode of presentation that Mary did not previously associate with the term "red."

The problem, as Thau sees it, is that there are very few candidate descriptions to fill the role required of them if we are to take redness to be identical to *that property*. He says, "the new belief … will have to be something like the belief *that fire engines have the intrinsic surface property of the thing she's seeing*," and that no "such belief can fully account for the content of the new belief that she registers …" (p. 192) Why not? On this point Thau is unconvincing, though making the case for this premise requires reworking the thought experiment yet again, a discussion of which would be beyond the scope of this review. It is essential that Thau be very convincing here, however, given the implausibility of the consequences of the argument, which are discussed below. Finally,

(7) Thus, redness is not identical to *that property*.

The use of redness in the above argument is arbitrary, so the argument applies *mutatis mutandis* to all of the standard cases of perceptible properties like more determinate shades of color (e.g. crimson), pitch, loudness, saltiness, and sweetness (p. 193): none of them are perceptually represented. Corresponding to those properties, however, are other properties or sets thereof that are perceptually represented, like *that property*, for example. What are these perceptually represented properties? The positive account that accompanies this negative claim is not articulated in enough detail to be satisfying, but in broad outline it proceeds as follows.

Given that colors are not perceptually represented, perceivers could be alike in their perceptually formed beliefs about the colors of objects while differing with respect to the properties they perceptually represent objects as having. So, the story goes, spectrum inversion is possible even though there is nothing more to perceptual awareness than the contents of representational states. Moreover, perceptual representation of properties is internally determined, whereas representation of properties like redness, squareness and the like is externally determined. This accommodates Thau's strong intuition that "*everything* about the way things visually seem is plausibly internally determined. Contrary to what some have claimed, I find it very hard to imagine that there could be *any* phenomenological difference between your physical duplicate and you" (p. 237). Thau believes, in addition, that there are good reasons for thinking that the content of perceptual beliefs about the colors, for example, is externally determined. His account therefore accommodates both the internalist intuitions about experience and the externalist intuitions about the contents of beliefs.

Now, in virtue of how their representation is internally determined, the perceptually represented properties are not nameable with public language predicates. What it is for these properties to be represented is for one to have a perceptual episode with them as their content, and the only sense in which one can think about them is by ostending them—"that property"—while one undergoes such a perceptual episode. It is essential to Thau's approach that the perceptually represented properties are unnameable, because if they were then one could substitute them into the argument we just reviewed and get the result that they are *not* perceptually represented just like colors, sounds, and smells (p. 223). Alternatively, of course, this requirement on unnameability suggests that there is a problem with Thau's argument.

Thau is quite clear that perceptual representations, like beliefs, are truth-apt, notwithstanding the fact that their contents are internally determined. He never mentions whether perceptual representations are generally veridical, however, and he never tells us what these perceptually represented properties are supposed to be. If they can be properties of the objects that seem to have them, then why can we only pick them out by having perceptual episodes with them as their contents? And if we can pick out those properties by some other means, then why can we not name them as we name properties such as redness? These are fundamental questions about Thau's view that his book leaves unanswered. It is also curious that whenever Thau tries to characterize the perceptually represented properties that correspond to the colors, they sound a lot like the colors themselves. For example (pp. 34–35), he says that they are the properties in virtue of which objects are intrinsically related to one another in much the same way that we judge the colors of objects to be intrinsically related to one another.

As mentioned above, one of Thau's motivations is the intuition that spectrum inversion is possible, which he traces to the intuition that the contents of perceptual representations are internally determined. Without a positive characterization of the perceptually represented properties, however, he has failed to establish the possibility of spectrum inversion. Inversion scenarios are those in which neither subject misrepresents the perceived objects even though their experiences of those objects differ in some significant respect. Shoemaker's view, from which Thau draws inspiration, is interesting because subjects can be inverted even though no one misrepresents anything *and* all that anyone is aware of is the content of mental states. Thau's view should do at least this much, but it seems as though for Thau inversion may be possible only at the price of misrepresentation. Subjects can be alike in judging green objects to be green, and yet differ in the corresponding properties that they perceptually represent green objects to have. But without knowing what those perceptually represented properties are it is impossible to know whether subjects can differ in this way without at least one of them perceptually misrepresenting green objects.

On the whole, this book is an interesting attempt to work out what the most plausible representationalist account of consciousness and cognition should look like. For Thau this involves a pure Millian account of mental representation coupled with some radical claims about perceptual representation. His version of the knowledge argument leads Thau to these claims, but absent a more fleshed out positive account of the nature of perceptually represented properties it is difficult to find his arguments convincing. On a more positive note, it is refreshing to see a new approach to the problems raised by spectrum inversion and the knowledge argument especially *vis-à-vis* the debates between Millians and Fregeans. The best way to find out whether the problems of consciousness are really a subset of the problems of mental representation is to develop a careful and fine-grained understanding of mental representation. Despite its limitations, *Consciousness and cognition* is an interesting contribution to that project.

John Kulvicki
Department of Philosophy
Washington University, St. Louis, USA
Email: jkulvick@artsci.wustl.edu

Duality of the mind: a bottom up approach toward cognition

Ron Sun
Mahwah, NJ: Erlbaum, 2002
ISBN 0805838805 (hardback, 296 pp., $59.95)

In his *Duality of the mind*, Ron Sun moves beyond the technical details of his work on human skill learning to develop a broad framework for studying human cognition. This background in skill learning is evident in his general approach to cognition. Sun's starting point is that cognition ought to be understood as emerging in its natural, daily context: human activity. Given this basis, Sun asks what the essential features of cognition thus conceived are, and how to capture these in computational terms. He lists first four behavioral characteristics:

(1) *reactivity*: a lot of human behavior consists of immediate and fixed responses to environmental features that do not require to be recomputed every time;
(2) *sequentiality*: human activities tend to consist of multiple steps that are temporally strung together;
(3) *routineness*: human activities often consist of habitual sequences of behavioral responses; and
(4) *trial-and-error adaptation*: learning reactive routines is essentially a trial-and-error adaptation process.

Together, these characteristics form the basis for learning skilled human activities that range from habitualized reactive routines to high-level intellectual skills. They also form the background against which Sun sets his more specific ideas concerning mind and cognition.

Most central in Sun's view is the claim that the mind is not a unitary system but consists of two separate parts, an implicit knowledge system and an explicit knowledge system. Differentiating between implicit and explicit knowledge is common, and has popped up in many slightly different guises such as procedural and declarative knowledge, sub-conceptual and conceptual processing, or non-conscious and conscious processing. However, Sun takes this distinction a step further. According to Sun, findings from skill learning and other data suggest strongly that there is a basic and qualitative difference between a low-level, connectionist system of procedural knowledge, the workings of which remain inaccessible to conscious scrutiny, and a high-level classic symbol processing system with consciously accessible declarative knowledge. To repeat, the mind is fundamentally *a dual system*.

Next to this central idea of the duality of mind comes the claim that the lower, implicit knowledge part is the primary one and that learning proceeds in a *bottom up* way. Sun opposes studies in which learning is to a large extent top down, and individuals learn first generic, verbal declarative knowledge that is only subsequently turned into specific procedural skills through practice. He holds that human skill learning ought not to require a large amount of *a priori* knowledge but is essentially based on bottom up derived implicit knowledge to which high-level explicit knowledge provides a later addition. The good thing of having two qualitatively different components to the mind is that this provides the option of working together in a *synergistic* way that generates better overall results. Sun mentions a final important characteristic of the mind—*modularity*. Within the two major systems of implicit and explicit processing, many other special purpose cognitive faculties might arise according to need or as a result of inborn tendencies.

Sun brings all these characteristics together in his own dual model of cognition, called Clarion, which consists of two levels. At the connectionist bottom level, reinforcement learning takes place that leads to the selection of appropriate actions for any sensory input. In addition there is a top level using classic symbols that executes a rule-learning algorithm. If some action, decided on by the bottom level, is successful, then the cognitive system extracts a rule that corresponds to this decision and adds the rule to the rule network. When necessary, the rule is subsequently revised. The top level is also capable of so-called plan extraction where explicit sequences of action steps are generated that do not require environmental feedback. It is here that the Clarion system moves beyond reactive actions that remain tied to the current environmental states and becomes capable of initiating and executing plans that require little or no immediate sensory feedback. Sun provides an elaborate appendix in which examples of the Clarion architecture dealing with mazes and minefield navigation are given. He also provides an impressive collection of human data that can be accounted for by simulations with his Clarion architecture.

So far, all this sounds like good, solid AI. Classic symbol processing and connectionist networks are brought together in a hybrid approach that incorporates the good characteristics of two different modeling techniques and also extends these characteristics by their synergetic interaction. However, Sun's ambitions and views go beyond an interest in a particular cognitive architecture as he links this architecture with more general, theoretical issues, drawing inspiration from philosophers and social scientists as Heidegger, Merleau-Ponty, Vygotsky, and Bourdieu. More specifically, he discusses Clarion in relation to situated cognition, consciousness, and sociocultural cognition. It is here that Clarion acquires a much more general meaning that goes beyond the architecture itself. However, this is also the place where Sun's story is sometimes less convincing.

Sun's discussion of symbol grounding and situated cognition deserves scrutiny in this respect. Symbol grounding is a well-known problem for AI. How are the symbols and representations used within an artificially intelligent system to be connected to the world outside, other then through an ultimately arbitrary interpretation by the model builder herself? A solution to this problem has been found by connecting symbols to sensory-motor processes and so giving the system its own interface with the external world. Its internal symbols then ought to reflect the external world in a non-arbitrary way. Sun develops this solution by stressing the importance, and the primacy of everyday activities. From Heidegger and Merleau-Ponty, he derives the idea that our existence in the world is fundamental, "That 'mindless' everyday activities, or coping with the world … is the basis of high-level conceptual thinking and intentionality" (p. 141). Sun uses the term "comportment" for this behavioral basis of everyday reactive and routinized activities that are direct and unmediated by high-level conceptual thinking. "Comportment" consists of a structural two-way connection, a pattern, between an agent and its environment. According to Sun, comportment constitutes the foundation of the cognition of agents. At this point, Sun explicitly moves away from standard views in artificial intelligence where internal knowledge is foundational, and places himself in the camp of situated cognition that stresses the environmental and bodily embeddedness of cognition.

Sun's embrace of situated cognition does not lead to downplaying the importance of representations in cognitive explanation as often occurs in this context. Comportment may be the foundation of mind, but it is not the whole structure. Sun identifies comportment with the bottom level of his Clarion model and stresses that symbolic thought occurs at the top level of Clarion as a different kind of processing. The top level builds upon and extends the reactive and routine activities of the connectionist bottom level with deliberate symbolic thought, while Sun also stresses the different processing capabilities of both levels. He views the mind as a hybrid, dual entity.

At this point one may want to criticize the move toward a dual and bottom up view of the mind. This is also the kind of criticism Sun seems to expect, and provides arguments and data against. On the other hand, one may as well ask whether Sun's move is going far enough. He claims that his model is in line with the general ideas of situated cognition. However, there are many differences. Situated cognition stresses the importance of bodily and situational factors for intelligent action. This must be taken literal, even when it leads to almost impossible demands on modeling. The specific motor and sensory characteristics must be taken into account, and so must the continuous reciprocal coupling between them, the physical environmental characteristics, and the timing constraints. Situated cognition stresses the physical behavioral interaction between an agent and its environment as the foundation of cognition. Sun agrees with this when he defends the foundational role of comportment; however, it is questionable whether the action selection mechanism of Clarion can bear this conceptual burden. While situated cognition (as well as comportment) centers around the issue of how actions are constituted by organism–environment interactions, in Clarion actions are no more than the result, or output, of action selection mechanisms at the top and bottom level. While situated cognition stresses the continuous sensory feedback of behavioral movements for cognitive functioning (e.g. O'Regan & Noë, 2001), Sun's Clarion only uses such feedback for the *learning* of unidirectional perception–action links. The discrepancy becomes highly visual in the figures. There Clarion is depicted as a feedforward system with arrows going into one direction. A situated approach will always stress the presence of a circular setup and invoke diagrams with arrows feeding back from the action to the perception side.

To conclude, Sun is verbally much more radical than his computational model of the dual mind warrants. While his conceptual story is founded on comportment, his explicit model remains firmly grounded in the skill learning literature and a relatively abstract, and classical, interpretation of human action. Action remains a matter of *deciding* what to do, rather than *doing* it. From a situated perspective, Sun's bottom level doesn't reach low enough to provide a foundation for cognition.

This discrepancy between Sun's claims about the general implications of his dual view and the limitations of the actual Clarion model can also be witnessed in the other two general topics dealt with by Sun, consciousness and social cognition. In his chapter on consciousness, he evaluates a large number of different cognitive models that, among other things, try to deal with consciousness by differentiating between two subsystems—just like Sun's own Clarion. He finds them all wanting for various reasons, except for Clarion. His main reason is that, in contrast to the other models described, Clarion offers two distinct *mechanisms* for the two levels that correspond to conscious accessibility, or not, for the two levels. Explicit symbol processes operating on localist representations at the top level provides a mechanism that explains conscious access of this level, while the distributed representations of the connectionist networks at the bottom level explains why processing at this level remain consciously inaccessible. It is unclear why Sun thinks that the explicit symbol processing in his Clarion model now suddenly provides an account of consciousness, while this has been one of the major difficulties in AI and cognitive science since their inception. Granted, Sun is at that point discussing access consciousness, which does not require the phenomenal aspects of qualia, but only a capacity for verbal report. However, he still ends with the claim that "In Clarion, qualia are the result of the two-level organization, as well as intricate structures involved in various fine-grained modules" (p. 189). Here, the book sounds more like a one-sided advertisement for Clarion than an even-handed discussion of the problems involved in consciousness.

The chapter on social and cultural aspects of cognition exhibits the same problem. Again there is an overly positive treatment of Clarion, now as a way of dealing with sociocultural factors. Clarion's bottom level is said to be capable of learning implicit social roles in a bottom up way. In addition, Clarion's top level provides also a way to assimilate explicit external knowledge within a social context, which can subsequently influence the bottom level in a top down fashion. Well, it may be possible to interpret Clarion in this suggestive and interesting way, but the symbol grounding problem will inevitably come banging on the door again. Clarion so far remains a very simple system that is not literally grounded in comportment—being involved in real, situated actions—and certainly not partaking in a social world. There remains a wide gap between Sun's philosophical story of comportment and the actual model.

Sun's pushing of the Clarion model may be overdoing it, but all in all this should not distract one from the many good things that must be said about the book. Sun's philosophical story places cognition firmly in the natural context from which it arises: human activity. His basic division of the mind in two different parts is a plausible way of dealing with empirical data that do not conform to the age-old conception of the mind as a unitary entity, his attention for sociocultural factors in cognition is laudable, and many of his more detailed observations and discussions are insightful and important. Next to this, the discussion of his Clarion model provides a good view of the state of computational work on cognition. That the philosophical and the model component of the book remain a duality is a blemish, but at least one that remains fully within the spirit of the *Duality of the mind*.

References

O'Regan, J.K. & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, *24*, 883–917.

Fred Keijzer
Faculty of Philosophy
University of Groningen
A-weg 30
9718 CW Groningen
The Netherlands
Email: f.keijzer@philos.rug.nl