'Bridge Out' on the Road to a Theory of Consciousness

Review of *The Conscious Mind: In Search of a Fundamental Theory* by David J. Chalmers

Gregory R. Mulhauser
University of Glasgow
Department of Philosophy
Glasgow G12 8QQ
U.K.

scarab@udcf.gla.ac.uk

PSYCHE, 2(34), October 1996
http://psyche.cs.monash.edu.au/v2/psyche-2-34-mulhauser.html

KEYWORDS: consciousness, dualism, epiphenomenalism, functionalism, panpsychism, supervenience, zombies.

REVIEW OF: *The Conscious Mind: In Search of a Fundamental Theory* by David J. Chalmers. Oxford University Press, $29.95 hbk., pp. xvii + 414, ISBN: 0-19-510553-2

1. Introduction

A clutch of '-isms' characterises the approach to consciousness which David Chalmers defends: dualism, epiphenomenalism, functionalism, anti-reductionism, and -- probably -- panpsychism. (The author would no doubt want 'naturalism' included in the list as well, but as we shall see, Chalmers' predilection to describe his theory as 'scientific' stretches credibility.) While the book does not, as far as I can see, move consciousness research significantly forward, Chalmers succeeds admirably in clarifying the philosophical terrain around and within each of these '-isms' and in questioning the usual assumptions which suggest some of them are mutually exclusive. Because nearly all of what follows is highly critical, I want to be explicit about one thing: I do *not* think this is a bad book. Throughout, most discussions keep to a very high standard; it's just that they include fatal flaws.

The book begins with Chalmers' most over-used rhetorical device, the motto 'take consciousness seriously'. This he elucidates (p. xii) as the assumptions that 1) consciousness exists and 2) it cannot be explained by explaining how either cognitive or behavioural functions are performed. The sound-bite returns with grating frequency

throughout the text, usually heralding a helpful reminder along the lines that if one takes consciousness seriously, then "the conclusions for which I am arguing must follow" (p. 110) or "property dualism is the only reasonable option" (p. 168). Quick to berate those who see merit in cognitive theories of consciousness, Chalmers does little better in simply defining them into irrelevance. Indeed, a charitable way to read the book is as a sustained attempt to rationalise the author's intuition on 2), without ever actually discharging the role the intuition plays as an assumption in his arguments. Indeed, he shows no concern to discharge it, proclaiming later that "it is almost impossible to argue *for* the premise... At best, one can try to clarify the issues in the hope that enlightenment sets in" (p. 168, emphasis original).

After an initial chapter of doing just that -- trying to clarify the issues and hoping enlightenment sets in -- Chalmers launches into an exposition of varieties of supervenience. Developing his analysis of the relationship between consciousness and the physical world in the framework of supervenience rather than 'identity' talk is a good move on Chalmers' part, although I worry that the long and tedious treatise in Chapter 2 -- by far the longest in the book -- may incline many readers only to skim it. While the discussion of supervenience adds little to the existing literature and the careful exploration of *a posteriori* necessity takes us no farther than Kripke (1972), wading through the chapter and grasping Chalmers' own use of the terminology is essential for making sense of much of the rest of the book.

In a nutshell, the important point for making sense of much of the rest of this review is that facts about consciousness might supervene either naturally or logically upon the supervenience base of physical facts. In the case of logical supervenience, fixing the supervenience base of physical facts and physical laws automatically fixes facts about consciousness. In the case of natural supervenience, by contrast, fixing all the physical facts and physical laws might still allow facts about consciousness to vary; in that case, we must add some extra natural laws into the supervenience base to fix the facts about consciousness. Chalmers believes consciousness supervenes only naturally on the physical world; that is, he believes consciousness cannot be explained without some new laws of Nature.

The only technical matter I find distracting in this respect is Chalmers' haphazard alternation between 'entailment' and 'implication' in describing the sense in which facts in the supervenience base fix supervening facts. The difficulty, which seems of little importance early in the book, returns later with Chalmers' poor discussion of the possibility that supervening facts might be necessarily true but unprovable with respect to the supervenience base (analogously to propositions which are true but unprovable with respect to a particular formal system). This relates to Chalmers' distinction between 'broadly logical possibility' and 'strictly logical possibility' to which he alludes on p. 35 but does not expand upon. In a footnote for that page, repeated almost verbatim on p. 52, he refers to "justifying" formal system axioms and rules according to some prior notion of logical necessity and possibility. The attempt at clarification only contributes to the suspicion -- underscored by later excursions on decidability and on combinatorial state

automata, to which we come later -- that the text is not informed by a clear understanding of formal systems.

2. Imagined Thought Experiments

Although Chalmers later claims to have "argued exhaustively for" (p. 184) the failure of consciousness to supervene logically on the physical world, in Chapter 3 he does nothing of the sort -- although he does repeat a host of thought experiments and arguments which have already appeared in the literature. His first example is the alleged logical possibility of a zombie; while zombies come in many flavours, Chalmers' is a particularly strong variety: a physical, functional, and *psychological* duplicate of himself who nonetheless has no phenomenal experience. Chalmers' zombie twin thinks, perceives, reflects on his own internal states, deliberates between chocolate chip mint ice cream and strawberry ice cream, and is even perceptually aware of a state called 'headache' when he eats his ice cream too quickly -- yet he experiences nothing at all. Throughout this and subsequent examples, Chalmers feigns complete bafflement at how there could be any *conceptual* connection whatsoever between the performance of cognitive functions (exemplified by the full-fledged psychology of his zombie twin) and the having of experience. This lack of connection he takes to illustrate the failure of facts about consciousness to supervene logically upon physical facts. Not until much later in the book do we discover what he may really believe: that (psychological) awareness is both necessary and -- wait for it... -- sufficient for (phenomenal) consciousness. But we come to that shortly.

(Of course, given that Chalmers has already *assumed* on p. xii that consciousness cannot be explained cognitively, it's hardly surprising that he should find no conceptual connection whatsoever. Indeed, there hardly seems any reason to go through any of these thought experiments selected from the literature if he has already *defined* the situation so. But, I shall try to read things sympathetically and pretend that Chalmers doesn't *really* mean to beg so many of the interesting questions about the relationship between cognition and consciousness. Of course, such a sympathetic reading does require rejecting many of those conditionals peppered throughout the text reminding us what must be the case if we 'take consciousness seriously' ...)

To the objection that we can't *really* imagine his zombie twin because we can't really imagine billions of neurons (or, I would add, even the full host of cognitive functions) in detail, Chalmers replies with the bold assertion that "we do not need to imagine each of the neurons to make the case...it is enough to imagine the system at a coarse level, and to make sure that we conceive it with appropriately sophisticated mechanisms of perception, categorization, high-bandwidth access to information contents, reportability, and the like" (p. 98). Not only does Chalmers think he can imagine his zombie twin, but he even imagines he knows precisely which coarse level features he needs for a good imagining! Not surprisingly, there's no argument here. I suppose we just have to take Chalmers' word for it, although I confess I personally find it extraordinarily taxing to try *imagining*, all at once, the full repertoire of cognitive capacities available even to myself, let alone to hypothetical zombie creatures. Dennett (1995), incidentally, comments insightfully on such purported acts of super-imagination.

The general form of the argument within which we find the super-imagination is this:

1. If David Chalmers cannot grasp any entailment relation between A and B, then there is no such relation.
2. David Chalmers cannot grasp any entailment relation between A and B.
3. Therefore, there is no entailment relation between A and B.

Taking A as the physical facts about the world and B as the facts about consciousness, the argument is roughly that given on pp. 102-103, and it is obviously valid. There is little evidence, however, that it is sound. Humility does surface briefly on p. 110, where Chalmers concedes that conceivability might be tied to the limits of human cognition; this admission comes in the context of *explanation*, however, where the relevance of such limits is manifest. Chalmers ignores the bearing of the point on intuitions about logical supervenience, where the relevance ought to be equally manifest.

Cognitive limitations enter again on pp. 138-140, where Chalmers attempts but fails to answer the objection that facts about consciousness might bear a relationship to the physical supervenience base analogous to the relationship between undecidable truths and the formal systems in terms of which they are phrased. The objection is significant because if there were a convincing analogy, it would suggest that we simply might not be able to grasp the logical supervenience relation even if there is one. Here, however, as elsewhere, a weak grasp of formal systems compromises the quality of the reply offered in the text. Indeed, an objector need not even appeal to the raw limits of absolute decidability -- the limits of computational tractability would do fine. Chalmers appears to view human cognizers as ideally rational beings for whom at least some limitations of the real world are irrelevant. Readers interested in a more realistic view of cognitive capacities will find a much more informed discussion in Cherniak (1984), although that article does not address questions specifically about supervenience relations.

After zombies, we're treated to inverted spectra and then to 'epistemic asymmetry' -- the assertion that even knowing all the physical facts about the world "would not lead one who had not experienced it directly to believe that there should be any *consciousness*" (p. 102, emphasis original). A curious double standard intrudes here, since in discussing logical supervenience Chalmers repeatedly suggests *only* that knowledge of A-properties should allow someone to work out the B-properties "*given* that they possess the B-concepts in question" (p. 70, emphasis added, and similarly on p. 36). Nothing in the text justifies this mismatch between the explication of logical supervenience and the argument against it for the case of consciousness.

An interesting feature which runs through this and the rest of the standard arguments and Chalmers' replies to the standard objections is the notion that "we find it conceivable that all these physical processes [in the brain] could take place in the absence of consciousness" (p. 110); only much later (pp. 178-179) does Chalmers discuss the notion of 'explanatory exclusion', attributing it to Kim (1989). This is the suggestion that we can in principle give a microphysical account of higher level physical processes without

explicitly mentioning higher level features, thus rendering things like pain, memory, and perhaps even consciousness irrelevant in a significant sense to explanations of human behaviour. Of course, goes the standard reply, many things, such as planets, dolphins, and Beatles memorabilia are explanatorily irrelevant in the sense that we can always in principle give microphysical accounts of their behaviour which make no mention of them. All these things logically supervene on the physical, however, so in another sense they 'inherit' explanatory relevance from the physical base upon which they supervene. Chalmers argues, however, that such problems carry more dire consequences for consciousness, precisely because of his conviction that consciousness *does not* logically supervene on the physical. (This all comes in the context of Chalmers' facing up to the epiphenomenalist nature of his approach, to which we come in a moment.) Yet the very intuition that explanatory exclusion is an issue underlies many of Chalmers' arguments for the failure of logical supervenience! *Of course* we can imagine "all these physical processes" in a brain without involving consciousness, just as we can imagine "all these physical processes" in a flying baseball without involving a flying baseball. I seriously doubt *anyone* -- not even a philosopher with super-imagination -- would suddenly conclude 'flying baseball!' if presented with a mountain of data about positions and momenta of every single particle in a baseball. Chalmers offers no reason to believe the two cases differ in a relevant way.

3. Dualism and Epiphenomenalism

After the zombies and related thought experiments preying on explanatory exclusion, Chalmers uses Chapter 4 to expound his brand of dualism and address the problem that it requires epiphenomenalism, which he believes might not be such a bad thing. He engages in some wild speculation about intrinsic phenomenal properties of matter, properties not even in principle open to scientific exploration of any kind, and about how we might be able to say experience is causally relevant if we stuff it far enough inside physical entities which *are* causally relevant to each other. (Not that this kind of 'causal relevance' actually makes any difference to the way the world works, of course...) Chalmers eagerly dubs his rendition of epiphenomenalist dualism 'naturalistic', but that seems a bit unfair to the language.

His critique of interactionist kinds of dualism (pp. 156-158), where Chalmers draws a false dichotomy between Eccles's psychon approach and theories which appeal to consciousness-caused wave function collapse, prompts a purely stylistic complaint. Not only is it difficult in this particular case to imagine how anyone who has read Eccles could have written what Chalmers has, but frequently in other places papers or books are cited with little critical evaluation or effort to link them substantially with the text. The overall effect is very often that citations seem like afterthoughts, bolted on to the text. Perhaps this is just an occasion of Chalmers' effective summarising skills obstructing a more effective exposition and discussion.

In any case, Chapter 5 carries on with an interesting problem with the epiphenomenalist dualist approach, which Chalmers calls the paradox of phenomenal judgement: the problem of explaining how it is that consciousness itself could be entirely irrelevant to

why we think we are conscious. Recall Chalmers' imagined zombie twin. He, too, *thinks* he's conscious and ruminates about *his* hypothetical zombie twin and regularly reflects on the ineffable feel of what it is like to be him. Chalmers wants us just to learn to be happy with all this (p. 184), instead of suspecting that something might be amiss in his reasoning.

Most interesting is Chalmers' floundering attempt to explain not just why we *judge* we are conscious -- which of course he must say is down to cognitive psychology -- but how we can *know* we're conscious. His only answer to the problem is to reject outright both causal and reliabilist theories of knowledge for the case of consciousness, his rejection riding on the back of what seems a rather overblown view of our knowledge of our own consciousness. Ultimately, Chalmers is faced with the bizarre notion that our conscious experience bears on what it is like to be us without actually having any bearing whatsoever on our psychology: "I *know* I am conscious, and the knowledge is based solely on my immediate experience. To say that the knowledge makes no difference to my psychological functioning is not to say the experience makes no difference to *me*" (p. 198, emphasis original). How anything can *matter* to Chalmers without its mattering psychologically is a puzzling question. Later he notes that "the intrinsic quality of the experience...plays no direct role in governing cognitive processes" (p. 207). Chalmers apparently is happy to relinquish any and all *conceptual* connection between knowledge and belief, allowing that the two might wander entirely independently from each other were it not for the lucky synchronisation achieved by his unexplicated special theory of knowledge.

Chalmers drives a similar wedge into the case of remembered experiences, where he also rejects a causal theory (pp. 200-201); his position amounts to the view that while the *cognitive* content of an experiential memory may be captured by a causal account, its phenomenal aspect cannot be. This unsatisfactory discussion finishes with the notion that "a causal connection to an experience is not required to remember that experience" (p. 201). It is here that we get the first indication Chalmers might believe a cognitive state (remembered) is sufficient for a phenomenal state (of remembering) -- but clearer indications come later.

In the end, the chapter's discussion leads nowhere except to an unspecified special epistemology of consciousness, a special epistemology to which Chalmers must appeal in salvaging his dualist view and which figures centrally as an unstated assumption in many of his earlier and subsequent arguments *for* the view. He does admit that he says little of substance about the special epistemology and suggests that "A full understanding of these issues would require a lengthy separate investigation" (p. 209). I would note only as a side observation that all these deep muddles into which the anchor of epiphenomenalism drags Chalmers are metaphysically transparent on a functionalist view; that is, while there might be immense *practical* difficulties in fleshing out the cognitive processes explaining all that we would like to explain about phenomenal judgement, such a project doesn't appear to involve any deep metaphysical quandaries about novel theories of knowledge.

4. Awareness and Consciousness

In Chapter 6, on the coherence between consciousness and the underlying cognitive processes, Chalmers begins to build a positive theory of consciousness; it is also here that the book really starts to come unglued. The first difficulty appears when Chalmers cavalierly brushes off the fact that a theory of consciousness of the sort he proposes is thoroughly untestable: "This worry will only come into play in a strong way if it turns out that there are two equally simple theories, both of which fit the data perfectly, and both of which meet the relevant plausibility constraints" (p. 217). It seems not to matter to Chalmers that his theory is untestable even *in principle*; of course this would be no problem if we were only after a *conceptual* theory -- but Chalmers makes it clear throughout the book that he's *not* after a conceptual linkage between consciousness and cognition, but a link governed by contingent natural laws above and beyond standard physics.

Given his interest in contingent laws of nature as opposed to conceptual truths, it is entertaining to observe Chalmers struggle to develop new natural laws contingently linking the physical and phenomenal, all the while appealing almost exclusively to *a priori* considerations. For instance, we get the 'detectability principle', which suggests that generally we have the capacity to form second-order judgements about our experiences: "Of course many experiences slip by without our paying any attention to them, but we usually have the *ability* to notice them: it would be an odd sort of experience that was unnoticeable by us in principle" (p. 219, emphasis original). I would have thought that such detectability is *constitutive* of conscious experience -- i.e., that a conscious experience which was not even in principle noticeable was no conscious experience at all. (But then, by Chalmers' lights, I don't take consciousness seriously anyway, since I don't define out of existence the possibility of a cognitive theory.) For epiphenomenalist Chalmers, however, the principle is a contingent natural law.

Referring to this and the closely related 'coherence principles', Chalmers audaciously asserts (pp. 233-242) that his principles "can play a central role in empirical work on conscious experience" (p. 233). But this is so much metaphysical hot air: all the work allegedly done by his principles is in fact done by the *assumptions* (such as that simple theories are preferable, that laws of nature apply equally across space and time, etc.) to which Chalmers appeals (pp. 216-217) in inferring his principles.

Later in the chapter, Chalmers finally lets us in on the little secret that "some kind of [psychological] awareness is *necessary* for consciousness" (p. 243, emphasis original), a statement which, from the context, seems clearly to mean the *logical* variety of necessity rather than mere natural necessity. (Certainly everywhere else in the book, such as for the whole of Chapter 7, Chalmers carefully flags any use of specifically natural necessity; there is little reason to think he's had a lapse here.)

Two pages later, we discover Chalmers thinks awareness is also sufficient for consciousness, *provided* we have on board his psychophysical 'bridging' laws. But the only psychophysical laws we've been offered up to this point in the text have come from

a priori considerations, together with the single first-person empirical observation that consciousness exists. (It is even clearer in the next chapter, where Chalmers argues for the 'principle of organisational invariance' -- functionalism, to the rest of us -- that the only bit of empirical information at work is this same observation that consciousness exists.) While many of the *a priori* considerations also take propositions about the character of existing physics as assumptions, recall that physical laws are included in the supervenience base relevant to questions of whether consciousness logically supervenes on the physical. Chalmers' position apparently reduces to the notion that awareness is necessary and sufficient for consciousness, provided consciousness exists. But what sort of sufficiency is that? It is bizarre to say that we know from *a priori* considerations that A is sufficient for occurrences of B, provided B exists at all. (We might instead say that A *together with* B's existence is sufficient for occurrences of B, but this pulls the teeth from the sufficiency.) The most straightforward conclusion is that Chalmers really does believe consciousness logically supervenes on the physical, given that logically supervening psychological awareness is both necessary and sufficient for consciousness. Alternatively, perhaps Chalmers is so smitten with the zombies he thinks he can imagine that he chooses to overlook the fact that nothing in his psychophysical laws is sufficiently empirical to save consciousness logically supervening on the physical.

The next chapter argues for the 'principle of organisational invariance', which, given Chalmers' belief that psychological awareness is entirely constituted by functional organisation, can be seen as nothing but a restatement of his coherence principles. Here, he specifically refers to functional organisation rather than awareness, claiming that "conscious experience arises from fine-grained functional organization" (p. 248), but this is really just extra mileage from the *a priori* considerations of the previous chapter.

Chalmers rehearses two thought experiments, both of which -- contrary to popular tale -- have appeared in one form or another elsewhere in the literature, in a bid to show that conscious experience naturally supervenes on functional organisation. The first is the gradual replacement scenario, and it takes the form of a *reductio* on the assumption that absent qualia are naturally possible. (A hypothetical organism with 'absent qualia' has no phenomenal experience -- no qualia -- whatsoever, although they might have a perfectly normal psychology.)

We imagine replacing neurons (or whatever) in the conscious subject with the analogous components from a functional isomorph (perhaps made of silicon, although it doesn't matter) who has absent qualia. The question then becomes whether qualia gradually fade out for the intermediate subjects, disappear abruptly at some percentage of replacement, or some more bizarre shift; Chalmers' preferred response is that none of these are acceptable possibilities and that the original assumption that a functional isomorph could lack qualia in the first place must be wrong.

What he apparently fails to grasp, however, is that nothing in the argument changes if we replace the assumption to be rejected with 'assume absent qualia are *logically* possible'. (And, obviously, if the argument were to succeed in showing that absent qualia are

logically impossible, it would also succeed in showing they are naturally impossible.) Chalmers' *only* appeal to any empirical premise in his argument (apart from the existence of consciousness associated with some functional system) is once again to a variant of his 'detectability principle': "In every case with which we are familiar, conscious beings are generally capable of forming accurate judgments about their experience, in the absence of distraction and irrationality" (p. 257).

Therefore, he concludes, it is entirely implausible that an intermediate victim of replacement therapy could be so systematically *wrong* about its own conscious experience, as would be required when, for instance, it continued psychologically to judge that it was having vivid experience of bright red, say, while only *really* experiencing faded pink. But of course his 'empirical' evidence is not empirical anyway -- for if it really were logically possible that our experiences could change significantly without our noticing, how could we possibly be led to believe this on the basis of our experience? Obviously, if our experiences changed significantly without our noticing, *we would not notice*! Chalmers' argument comes down to *a priori* considerations alone: if it goes anywhere, the argument is far more on the side of rejecting the *logical* possibility of faded qualia and accepting the logical supervenience of consciousness on the physical.

Indeed, Chalmers' own concession that awareness is necessary for consciousness is at odds with his feigned open-mindedness about the logical possibility of such faded qualia: speaking of his faded-qualia functional isomorph Joe, Chalmers says "Joe sees a faded pink where I see bright red, with many distinctions between shades of my experience no longer present in shades of his experience" (p. 256). On the very next page, Chalmers reminds us, "To be sure, fading qualia are *logically* possible" (p. 257, emphasis original) -- yet Joe's phenomenal experience of a single shade of pink when presented with a collection of red shades which Chalmers distinguishes is *not* supported by the *necessary* awareness of a single shade of pink. It is instead matched with Chalmers' own awareness of multiple shades.

The second thought experiment targets inverted qualia. This time we imagine a switching system which allows us to flip between having some portion of the subject's normal processing done either by its brain or by the corresponding components of a functional isomorph with inverted qualia. The result is that as the switch is flipped, the subject's conscious experience *changes*, but since functional organisation remains constant, psychological processes remain the same: while the subject's colour experiences can be made to 'dance' before its eyes, it doesn't even notice.

Once again, Chalmers says this is logically possible, although he admits "the case is so extreme that it seems *only just* logically possible" (p. 269, emphasis original). But of course for someone who claims there is no conceptual connection between consciousness and cognition, there ought to be no difficulty whatsoever with imagining this scenario or even more outlandish ones. Why couldn't Mike Tyson be living the phenomenal life of Mother Theresa, for instance? For someone with Chalmers' powers of imagination, there ought to be no conceptual roadblock at all to imagining Tyson-style psychology together with Mother Theresa-phenomenology. It oughtn't to be "only just" logically possible, it

ought to be downright obvious. On the other hand, of course, Chalmers has also confessed that awareness is necessary for consciousness, in which case he's just plain wrong about dancing qualia: they're logically impossible.

As before, Chalmers' single 'empirical' premise, apart from the existence of conscious experience associated with some functional system, is his detectability principle, that "when one's experiences change significantly, one can notice the change" (p. 270). But, yet again, this is not *really* an empirically-motivated premise, because its failure is not empirically detectable in principle.

Chalmers addresses such concerns (pp. 274-275), but incompletely. He does charitably concede that "Some might dispute the logical possibility of...[fading or dancing qualia]...perhaps holding that it is constitutive of qualia that we can notice differences in them" (p. 274). He is right to caution that this view leads only to "the logical necessity of the *conditional*: if one system with fine-grained functional organization *F* has a certain sort of conscious experience, then any system with organization *F* has those experiences" (p. 274, emphasis original). But this leads also to the conclusion that a very significant portion of the book, allegedly dedicated to establishing contingent psycho-physical laws, has really established only that if consciousness exists at all, then this is the relationship it must bear to the physical. Questions about Mike Tyson with Mother Theresa phenomenology are all to be settled with *a priori* considerations; the only remaining question is whether Mother Theresa *has* any phenomenology.

Chalmers' survey of the available arguments puts in place a large and robust framework for understanding how conscious experience relates to the physical world -- all on the basis of *a priori* considerations -- and the single sticking point is just whether there is any conscious experience. But the relationship Chalmers describes is a very peculiar one indeed: it suggests as before that some set of propositions (about the physical world) is sufficient for some other propositions (about consciousness), *provided* consciousness exists at all. It is as if someone were to argue that ten plus ten plus ten equals thirty, *provided* that any thirties exist. The zombie intuition, the idea that we could subtract off the conscious experience whilst leaving all the rest of this robust framework unaffected, is all that remains to Chalmers' project to build a theory of consciousness. Super-imagination looks more suspect than ever. (With the confusion about 'contingent' psychophysical bridging laws derived from *a priori* considerations cleared away, incidentally, Chalmers' position looks remarkably similar to 'metaphysical supervenience' grounded in Kripkean *a posteriori* necessity -- an approach of which he is openly critical.)

5. Information, Functionalism, Diminishing Returns

From this point in the book onwards, it becomes progressively more difficult to offer positive commentary. Chapter 8, called "Consciousness and Information: Some Speculation" is, not surprisingly, speculation. Throughout, Chalmers uses his rendition of Shannon's framework, making no mention of the more modern approaches to information, such as Greg Chaitin's algorithmic information theory (1987), which have

emerged in the decades since Shannon and Weaver's 1949 classic. (For an information theoretic approach to mind based on the newer work, see Mulhauser 1997.) His comment (p. 280) on the remote relationship between his own view of information and Dretske's (1981) is at odds with what both actually write. Chalmers, like Dretske, bases his view of information on Shannon. Chalmers is right that Dretske is interested in semantic notions of information but wrong that his own view as expressed in the book, despite being based on the very same mathematical notions, somehow yields a notion of information which isn't *about* anything. Although he ignores the fact, Chalmers' information is information *about* the causal processes with respect to which he defines it.

Later, Chalmers suggests the structure of phenomenology is actually isomorphic to the information theoretic structure of a functional system giving rise to the phenomenology. But apart from the fact that this is obviously false -- consider subliminal advertising, for instance, the very point of which is to affect later behaviour *without* affecting phenomenal experience, indicating that the subject's sensitivity to information in a particular experience is *not* matched by the details of their experience -- the suggestion rests on a mathematically hopeless notion of what it means to instantiate information physically. (This comes to a head in the next chapter.) The chapter continues with bizarre explorations of panpsychism, a side effect of one way Chalmers might link information spaces and phenomenal ones, according to which "there is experience wherever there is causal interaction" (p. 298).

Most of the following chapter, a defence of 'strong AI', rests on Chalmers' failure to grasp the vacuity of naive functional theorising (identifying psychological functions with abstract machines), as explored in different ways by Block (1978), P.M. Churchland (1981), or Putnam (1988), for instance. Chalmers' own criterion for implementing a combinatorial state automaton (p. 318), or CSA, is grossly inadequate, and despite his protestations to the contrary (p. 319 and back on p. 252), it does nothing to circumvent the standard criticisms of naive functionalism. (Although Chalmers chooses to ignore the standard troubles with machine functionalism, they concern other authors enough that an entire class of teleofunctionalist approaches has emerged in response; see Dennett 1975, Bogen 1981, Lycan 1981 and 1987, Millikan 1984, Papineau 1987, and Sterelny 1990.) The most significant problem is that Chalmers' requirement that there be an injective mapping from physical states to automaton states is trivially satisfied (and stipulating a surjection raises other difficulties). Perhaps the most blatant indication that Chalmers has not examined the deep issues here comes with his assertions (pp. 318 and 320) that a CSA description captures a system's causal organisation. That will be news to all those silly philosophers of science who have struggled with the problem of un-trivialising descriptions of causal organisation at least since the inception of the 'covering law' model in Hempel and Oppenheim's seminal 1948 paper; adding insult to injury, Chalmers' blithe appeal to counterfactuals in constraining CSA implementation ignores a snarl of difficulties first explored half a century ago by Chisholm (1946), Goodman (1947), Kneale (1950), and subsequent authors.

Worse, Chalmers apparently thinks his notion of implementing a CSA is equivalent to implementing a functional system. He uses this idea to argue for strong AI by saying that

"for a given conscious system *M*, its fine-grained functional organization can be abstracted into a CSA *M*, such that any system that implements *M* will realize the same functional organization, and will therefore have conscious experiences qualitatively indistinguishable from those of the original system" (p. 321, emphasis original). Yet, pretending for a moment that Chalmers' definition for implementing a CSA is not *entirely* mathematically trivial, it remains that infinitely many functional organisations may implement the same CSA: a system based on a giant Block-style lookup table, for instance, can be gerrymandered to satisfy any desired CSA specification, yet its internal causal organisation may be completely different to that of other functional systems Chalmers takes to implement the same CSA. Chalmers gives no reason to believe lookup tables are subjects of conscious experience.

Later in the same chapter, Chalmers replies inadequately to objections from incompleteness. As in the case of quantum theories of mind, I agree wholeheartedly with Chalmers that incompleteness is utterly irrelevant to the task at hand -- but, as before, I feel it is only appropriate that such topics be discussed in an informed and careful manner if they're to be brought up at all. It is remarkable to me that so many philosophers, Chalmers included, profess to address incompleteness seriously (!), yet so few have apparently taken the time to explore related developments from the six decades since Goedel's 1931 paper. For instance, Greg Chaitin (1987), mentioned previously, has generalised incompleteness to its most universal form, hugely simplifying things along the way. (That a clear grasp of Chaitin's approach renders patently obvious the irrelevance of incompleteness to questions about minds as formal systems perhaps goes some way toward explaining why half the philosophers ignore it -- but not so for the other half.) Nothing in the discussion is either new or especially clear. Chalmers points to his own 1995 book review for further details of his approach to the topic, but a reading of that article only strengthens the suspicion that his case would have been stronger with the whole issue omitted.

Immediately after that discussion, Chalmers has a quick go at debunking objections that discrete and continuous systems might have different computational powers (pp. 330-331). His point is that continuous systems would need to exploit infinite precision to exceed the powers of discrete systems. Readers interested in an analogue system which computes in polynomial time a superset of the Turing-computable functions and which does so with finite linear precision will find it in Siegelmann and Sontag (1994).

The final chapter addresses the interpretation of quantum mechanics (which is, like incompleteness, another popular topic for contemporary philosophising). I believe readers will find it far more worthwhile to delve into some real quantum mechanics, such as Omnes (1994). The upshot of Chalmers' discussion is a strong preference for the Everett interpretation, according to which the entire cosmos is literally in a state of quantum linear superposition.

He argues audaciously (pp. 349-350) that his theory of consciousness independently predicts results which the Everett interpretation requires to account for the fact that we only ever experience one branch of the Everett universes, thus giving us "a powerful

argument for" -- you guessed it! -- "taking the Everett interpretation seriously" (p. 351). What Chalmers fails to point out is that plain old materialist functionalism (or almost any mainstream materialist theory of mind) *also* yields the right perspectival fix on the many worlds approach. (Perhaps this is why physicists didn't immediately chuck out the Everett interpretation a few decades before Chalmers came along to rescue it with his theory.) Does that mean all materialist functionalists should take the Everett interpretation above all others? I think not.

Overall, those who look to this book in hopes of finding fresh leads in the quest to explain consciousness are apt to find only disappointment. The widespread publicity which surrounded *The Conscious Mind* even before it was available in print, suggesting it would contain startling new arguments and thought experiments, turns out to be overly optimistic. Perhaps it would have been difficult for any book to fulfil such expectations; but as a broad survey of the field, harbouring comparatively few outright errors, the book is a fine success. Its greatest use may be for those unfamiliar with the philosophical literature, but it will also challenge the thinking of even the most self-assured aficionados. Although it comes nowhere near its hoped-for destination of a believable dualist, epiphenomenalist, functionalist theory of mind, in this case it is the journey which matters. It is a journey which I found undeniably educational, whatever my personal quarrels with the text, and which would no doubt be similarly stimulating for the bulk of researchers presently puzzling over the challenge of consciousness.

References

Block, N. (1978) 'Troubles with Functionalism', in Savage (1978).

Bogen, J. (1981) 'Agony in the Schools', *Canadian Journal of Philosophy* 11: 1-21.

Chaitin, G.J. (1987) *Information Randomness & Incompleteness: Papers on Algorithmic Information Theory*. Singapore: World Scientific.

Chalmers, D.J. (1995) 'Minds, Machines, and Mathematics', *PSYCHE* 2:1.

Cherniak, C. (1984) 'Computational Complexity and the Universal Acceptance of Logic', *Journal of Philosophy* 81: 739-55.

Chisholm, R.M. (1946) 'The Contrary-to-fact Conditional', *Mind* 55: 289-307.

Churchland, P.M. (1981) 'Eliminative Materialism and the Propositional Attitudes', *Journal of Philosophy* 78: 67-90.

Dennett, D.C. (1975) 'Why the Law of Effect Will Not Go Away', *Journal of the Theory of Social Behaviour* 5: 179-87.

Dennett, D.C. (1995) 'The Unimagined Preposterousness of Zombies', *Journal of Consciousness Studies* 2(4): 322-26.

Davidson, D. and G. Harman, eds. (1972) *Semantics of Natural Language*. Dordrecht: Reidel.

Dretske, F.I. (1981) *Knowledge and the Flow of Information*. Cambridge, Massachusetts: MIT Press.

Goodman, N. (1947) 'The Problem of Counterfactual Conditionals', *Journal of Philosophy* 44: 113-28.

Hempel, C.G. and P. Oppenheim (1948) 'Studies in the Logic of Explanation', *Philosophy of Science* 15: 135-75

Kim, J. (1989) 'Mechanism, Purpose, and Explanatory Exclusion', *Philosophical Perspectives* 3: 77-108.

Kneale, W. (1950) 'Natural Laws and Contrary-to-Fact Conditionals', *Analysis* 10: 123.

Kripke, S. (1972) 'Naming and Necessity', in Davidson and Harman (1972).

Lycan, W.G. (1981) 'Form, Function, and Feel', *Journal of Philosophy* 78: 24-49.

Lycan, W.G. (1987) *Consciousness*. Cambridge, Massachusetts: MIT Press.

Millikan, R.G. (1984) *Language, Thought and Other Biological Categories*. Cambridge, Massachusetts: MIT Press.

Mulhauser, G.R. (1997) *Mind Out of Matter: Topics in the Physical Foundations of Consciousness and Cognition*. Dordrecht: Kluwer Academic Publishers.

Omnes, R. (1994) *The Interpretation of Quantum Mechanics*. Princeton: Princeton University Press.

Papineau, D. (1987) *Reality and Representation.* Oxford: Blackwell.

Putnam, H. (1988) *Representation and Reality*. Cambridge, Massachusetts: MIT Press.

Savage, C.W., ed. (1978) *Perception and Cognition: Issues in the Foundations of Psychology*. Minnesota Studies in the Philosophy of Science, vol. 9. Minneapolis: University of Minnesota Press.

Shannon, C.E. and W. Weaver. (1949) *The Mathematical Theory of Communication*. Urbana, Illinois: University of Illinois Press.

Siegelmann, H.T. and E.D. Sontag. (1994) 'Analog computation via neural networks', *Theoretical Computer Science* 131: 331-60.

Sterelny, K. (1990) *The Representational Theory of Mind*. Oxford: Basil Blackwell.