

Placing Qualia in the Head

Review of *Locating Consciousness* by Valerie Gray Hardcastle

Derek Browne

Philosophy Department
University of Canterbury
Private Bag 4800
Christchurch
New Zealand

d.browne@phil.canterbury.ac.nz

Copyright (c) by Derek Browne

PSYCHE, 3(1), February 1997

<http://psyche.cs.monash.edu.au/v3/psyche-3-01-browne.html>

KEYWORDS: consciousness, memory, neuropsychology, qualia.

REVIEW OF: Valerie Gray Hardcastle (1995) *Locating Consciousness*. Amsterdam/Philadelphia: John Benjamins Publishing Company (Advances in Consciousness Research, Volume 4). Pbk. pp. xviii + 264. ISBN 1-55619-184-7 (US). Hfl 70 pbk.

It is a foundational principle of the cognitive sciences that the function of brains is to process information in order to produce adaptive behaviour. One reason why this is hard to dispute is that the notion of an information processing system is so general that it can accommodate just about any plausible theory of mind. Even so relaxed a framework, however, is problematic where consciousness is concerned: it is not at all obvious whether, and if so how, consciousness contributes to the information processing functions of the brain. This is the conundrum that motivates Valerie Gray Hardcastle (Philosophy, Virginia Polytechnic Institute and State University) in her book, 'Locating Consciousness.' She offers her work as "an extended example of interdisciplinary research" in which philosophical arguments are supported with data from "cognitive and developmental psychology, AI programming, linguistics, clinical neurology, neurophysiology, and neuropsychology" (p. xv). She proves to be a knowledgeable guide to the relevant science. The more philosophical excursions in which empirical data are used to illuminate the nature of consciousness are interesting enough but not always convincing, for a variety of reasons.

Hardcastle aims to "examine what neuropsychology can tell philosophers about qualitative experience" (p. xiv). She thinks that bottom-up strategies of inquiry are more

likely than top-down strategies to bear fruit at present (p. 171). She argues for a replacement of the usual functionalist stance by a more 'structural' approach which gives greater weight to the neurophysiological underpinnings of consciousness. This doesn't really mean that consciousness is not a functional kind of thing. It just means that consciousness is to be explained by events at levels below the top-most level of psychological events.

The contributions that science can make to understanding conscious experience are grouped within the answers to three questions (p. xiv):

1. What are the appropriate properties of the mind and the brain to study in order to develop a theory of consciousness?
2. What informational role does consciousness play in our psychological life?
3. How does the underlying neurophysiological structure of consciousness relate to higher-level information processing descriptions of consciousness?

Her principal conclusion is that the theory of multiple memory systems offers promise of a scientific theory of consciousness. Specifically, contents are conscious if they are currently activated in semantic memory.

The plausibility of any neurophysiological theory of consciousness will depend in part on the detailed inventory of contributions made by the brain sciences. But it will also depend on what one accepts as facts about consciousness that any such theory must explain. Hardcastle thinks that the phenomenon to be explained is usefully identified as a type of state such that there is always something it is like to be in tokens of that state. Apparently, Nagel-speak resonates with significance for her. It does little for some of us. So I don't think it should be assumed without further investigation that Nagel-speak succeeds in picking out a well-defined class of phenomena.

Many philosophers have been inspired by Nagel (1974) and friends to think and talk about consciousness in a proprietary vocabulary which (a) all but guarantees the intractability of consciousness as a subject for science, and (b) doesn't even fit the phenomenology of experience all that well. It's a shame Hardcastle didn't scrutinize the familiar rhetoric before repeating it. Here she is opening Chapter One: "That we have minds is a wonderfully eerie fact about us. But minds are strange indeed, for they are conscious - at least in part. We have astoundingly vivid perceptions of the world. I go to the symphony and hear symbols crashing, flutes warbling, violins sighing, tubas booming. I see the conductor waving her hands, the musicians concentrating, patrons shifting in their seats, and the curtains gently and ever-so-slightly waving. I smell the perfume of the woman next to me, the damp musk of the chairs, the ink on the program. What is a mind such that it has these amazing powers?" What this passage tells us is that, in sense experience, we perceive a richly qualified world. So the world contains conductors and crashes and ink: what's eerie about that? It only begins to look eerie if you infer straight off that this rich panoply of properties is internal to the mind, and if thereby, in a single, sensational elision, you erase the distinction between mind and world. The phenomenology of perception presents the qualities of which we are aware -

the colours of surfaces, the sounds of musical instruments, the scent of perfume - as distal. This is robust phenomenology. It is possible that careful philosophical argument could unsettle the appearances and show that sense experience actually makes us aware (only) of things and properties that are inside our heads. But let's not assume this in advance of argument, and let's not pretend that internalism about mental contents gets any obvious support from phenomenology. (See Dretske, 1995.)

Fortunately, the harm done by stepping off on the wrong foot is limited. By the time Hardcastle has reached p.133, she is ready to describe qualia as "introspectively slippery and intellectually suspect."

A substantial part of *Locating Consciousness* reports scientific findings and conjectures that are useful in addressing the many problems of consciousness. In general, Hardcastle's evident grasp of the relevant sciences serves to limit the influence on her of many of the bad ideas and strategies of inquiry that circulate widely in contemporary philosophy. Thus my trepidation at embarking on yet another foray into inverted spectra was quickly stilled by the discovery that Hardcastle does not practice the intuitive method that still blights so much philosophy of mind. Instead, she spends time on Larry Hardin's (1988) argument from the neuropsychology of human colour vision to the untenability of inverted spectra hypotheses. It's a nice argument, one which rests, not on the shifting sands of a priori intuition and thought experiments about remote possible worlds, but on the firmer ground of empirical science and, specifically, on the opponent processing theory of colour vision. This beautiful theory is usually presented as the best theory of colour vision we have. Unfortunately (says Hardcastle, p. 31), there is "precious little independent evidence" supporting it. For a philosopher who is unabashedly naturalistic and is writing a book which extols the capacity of science to illuminate philosophical issues, this is admirably frank. (There is of course no comfort in this result for intuitionist philosophy, and Hardcastle certainly does not suggest otherwise.)

"What are the major psychological and neurophysiological differences between conscious states and unconscious ones?" asks Hardcastle (p. 57). That is, what differences are there apart from the difference that the former are conscious and the latter are not? An answer to this question would enable us at least to predict the appearance of consciousness in a cognitive system. It probably would not enable us (she admits, p. 83) to explain all the experiential differences among qualia. The key idea is that consciousness is activity in a semantic memory system. Here 'memory' is construed broadly as involving not only the storage of previous perceptions but also the processing of current ones. That is, memory systems are responsible for the processes through which perceptual inputs are interpreted according to stored information.

Memory researchers disagree over how many different kinds of memory there are. Hardcastle adopts the position of Tulving and Schacter (1990), who identify four distinct memory systems: procedural, semantic, episodic and an implicit memory system that Hardcastle calls 'structural' memory. Structural memory is activated in priming events. It is distinct from the memory system that handles semantic information. It's unclear to me, however, whether the distinction between a memory system that is implicated in priming

events, structural memory, and a semantic memory system, adequately accommodates the evidence for semantic priming. Perhaps Hardcastle's view is that the experimental evidence does not unequivocally show that semantic information is stored in priming events (p. 204, n112); perhaps her view is that the structural memory system which is activated by priming events has access to some semantic information, but only in a very shallow way; perhaps, as I will shortly suggest, the distinction between 'syntax' and 'semantics' is not a sharp one.

However that may be, contents in structural and semantic memory systems are accessed in different ways, as is illustrated by the finding that blindsight patients can extract 'syntactic' but not 'semantic' information from their blind fields. These patients can typically recognize letter shapes but cannot read words - as evidenced not by spontaneous judgements but by answers to forced-choice questions. Hardcastle's interpretation of this result is that the structural memory system has access to the information in blindfields and can pass it to motor systems which guide answers to forced-choice questions, whereas the semantic memory system has no access to blindfield contents, with the consequence that there is no conscious experience of those contents.

Hardcastle's major hypothesis (p. 85) is that contents are conscious if they are activated in the semantic memory system. This is a bold move. Her evidence is that in certain kinds of experimental setups, the processing of structural information is separable from the processing of semantic information, and that only the latter kind of information is consciously available. But it's a large step from here to the conclusions that (i) activated contents in semantic memory are always available to the kinds of probes that define the idea of introspective accessibility, whereas (ii) activated contents in structural memory are not (ever) available to such probes. I'd very much like to see the evaluation of a wider range of evidence here. I'm also unsure of the distinction between 'structural' (or 'syntactic') and 'semantic' interpretations of signals. A lot of information can be built into 'structure' (Millikan, 1993). It also seems to be the case that some pretty sophisticated recognition processes can occur and can have effects both on other realms of cognition and on behaviour without making it through to consciousness - at least, as evidenced by subsequent probes. Is it really apt to describe all such processing as nonsemantic? Perhaps Hardcastle's conjecture is right and consciousness is located in just one of two neurally discrete systems. Yet it might also be true that the functional differences between these two systems are not aptly captured by the classical distinction between syntax and semantics. Is there really a difference of kind in the functioning of the two systems or just, as Daniel Dennett (1995) would say, a difference in depth of processing?

Hardcastle concedes (p. 93) that her hypothesis doesn't explain why this type of interpretative process should give us consciousness, but she thinks that progress can be made on this 'hard question' (to use the dull name that is gaining unaccountable popularity). She begins with 'Marr's paradox' (p. 104): none of the representations occurring at any of Marr's demarcated stages of visual processing correspond to what we consciously experience in the visual domain. What does 'correspond' mean? Astonishingly, Hardcastle reads it as 'resembles': nothing in our conscious experience resembles the computational or brain states that instantiate them. But there is no good

reason to think that representational vehicles should resemble their contents (this token of the word 'red' isn't red). I'm not impressed by a 'paradox' that depends on ignoring the basic distinction between representational vehicle and representational content. Perhaps the paradox is supposed to be generated by nonrepresentational qualities of mental states: the brain sciences haven't found anything that resembles or looks like feelings or other qualia. To evaluate this idea, however, we need first to work out what feelings or their neural instantiations should look like. Hardcastle seems to recognize that the 'paradox' is not well-formed but she doesn't labour to produce a more compelling version. This is a pity. Is it really necessary for the modern representational theory of mind to recapitulate the history of Locke's theory of ideas? Locke was a great philosopher, and if we can see further than he did it is because we can stand on his shoulders.

Hardcastle turns to the problem of unified perception, the binding problem. (The narrative links in her book are not always transparent.) She is critical of the famous 40Hz hypothesis, and sketches instead an account in terms of 'higher order patterns of bifurcation in an attractor phase space.' This is bracing stuff, but so far as I can see, it doesn't address 'Marr's paradox' or 'the hard problem' at all. At most, data on cortical maps suggest hypotheses about the localization of perceptual consciousness. This is not a negligible result: the book is, after all, called *Locating Consciousness*. But it doesn't begin to address philosophically familiar issues about the relation (identity? implementation? supervenience? constitution?) between events of consciousness and neural events in those locations.

Hardcastle claims next that the empirical hypotheses about semantic memory shed light on philosophical concerns about absent qualia. I don't think this is really true. She recounts some of Sydney Shoemaker's (1984) arguments against absent qualia hypotheses: in brief, given that introspective judgements are causally based, any creature that is a functional isomorph of you must be introspectively identical to you. Hardcastle's discussion contains nothing that will be unfamiliar to philosophers working on this topic, while those to whom these philosophical issues are unfamiliar would be better advised to read some of Shoemaker's fine articles instead. How does Hardcastle think that her empirical hypothesis - that conscious events are events in semantic memory - illuminate these problems? So far as I can see, the point is that theoretical definitions of qualitative psychological states will include not only 'horizontal' causal relations between psychological states (of the sort pictured in information processing flow charts) but also 'vertical' causal relations between those qualitative states and the neural states that implement them. If I and my metaphysical twin are neurophysiologically identical, then it is not possible that I have any qualitative psychological states which are absent in him. This is a plausible position for a physicalist to take, but the multiple memory systems hypothesis contributes nothing that is distinctive. Indeed, as Hardcastle recognizes (p. 148), there is for physicalists a fast argument to the same conclusion: if you accept that mental differences supervene on physical differences, then any mental difference (including introspective differences) between real pain and ersatz pain must supervene on physical differences. Her real concern, I think, is to combat the classical functionalist stance that takes cognitive science to be implementation-neutral. May the force be with her.

Hardcastle's principal hypothesis is "that conscious phenomena is (sic) aligned with the activation of memories in our semantic memory system" (p. 151). How should we evaluate this hypothesis in light of the current vogue for 'executive processing' models in cognitive psychology? Since semantic memory provides inputs to executive systems (global workspaces, etc.), the prima facie difference is that for Hardcastle, conscious experience occurs upstream of executive processing. Hardcastle's defense of her own hypothesis against executive processing models has, she says (p. 152), the following interesting consequence: consciousness is probably not a phenomenon that occurs at higher, psychologically described levels of cognition, the level at which, for instance, executive functions occur.

Suppose we grant (she says, p. 161) that the prefrontal cortex implements a supervisory system which is involved in higher level control in novel situations, in planning, and so on. Still, the identification of consciousness with the executive system fails because the data also establish that this system is "divorced from phenomenal experience." Prefrontal lesions typically produce a deterioration in cognitive processes (consistent with the executive system hypothesis) but no degradation of sensory experiences. Indeed (p. 162), patients can apparently lose their entire frontal cortex and still be conscious. The data on which executive system theorists rely is also consistent with Hardcastle's hypothesis that conscious contents form a subset of inputs to the executive system.

Hardcastle does not see herself as offering an alternative model on the same level of analysis as executive processing models. She says (p. 170): "We need to do a better job of predicting the occurrence of qualitative experience and localizing it in the brain before we can begin to construct hypotheses about the purposes behind the experiences. We really don't know enough neuroscience to attempt this sort of explicit reduction of conscious experience or to rely on neurophysiological data to justify the higher level computational theories of cognition. The lesson here is that instead of fairly sweeping executive theories, one should pursue a theory of consciousness that tries to locate consciousness in the brain and to outline the processes of that particular spot (or the relevant interconnected circuits) before attempting to explain how those functions fit into our information processing picture of the mind." The resulting theory would be "more 'bottom-up' than 'top-down'."

Perhaps. But reasoning about information processing in the brain plays an important role in triangulating conscious experience. Hardcastle seems to impose few conditions on the acceptance of introspective reports. Yet these reports are themselves outputs of an information processing system. Perhaps they can sometimes be taken, at face value, as noninferential reports of the qualities of conscious experience. But equally there are numerous situations in which alternative hypotheses about the aetiology of such reports are preferable, hypotheses which don't accept them at face value. So at the very least, Hardcastle should acknowledge a role for cognitive models which describe the processes that produce first person reports of the contents of consciousness.

Valerie Hardcastle displays a greater competence in neuroscience and cognitive psychology than most philosophers of mind can muster. This makes her work valuable

for philosophers who want to learn some relevant science. The more philosophical parts of the book (on methodologies for cognitive science, puzzles about qualia, and so on) are less successful. I cannot really see in this book even a sketch of an answer to 'the hard question,' or any hint of a way in which the neuropsychological hypotheses might close 'the explanatory gap.' Still, I think her central thesis, which connects consciousness to the activation of contents in semantic memory, was well worth stating, and deserves further consideration.

References

- Dennett, D. (1995). The path not taken. *Behavioral and Brain Sciences*, 18, 252-3.
- Dretske, F.(1995). *Naturalizing the mind*. Cambridge, MA.: Bradford/MIT.
- Hardin, C. L. (1988). *Color for philosophers: Unweaving the rainbow*. Indianapolis: Hackett.
- Millikan, R. G. (1993). On mentalese orthography. In Dahlbom, Bo (Ed.) *Dennett and his critics*. Oxford: Blackwell.
- Nagel, T. (1974). What is it like to be a bat? *Philosophical Review*, 83, 435-451.
- Shoemaker, S. (1984). *Identity, cause, and mind*. Cambridge: Cambridge University Press.
- Tulving, E. & Schacter, D.L. (1990). Priming and human memory systems. *Science*, 247, 301-306.