Neurophilosophy comes of age?

Brain-Wise: Studies in Neurophilosophy

Patricia Smith Churchland

Cambridge, Mass.: MIT Press, 2002. 471 + xii pp. ISBN 0-262-03301-1 $65.00 hb;

0-262-53200-X $25.00 pb.

Review by Dan Ryder

Nearly twenty years ago, Patricia Churchland wrote Neurophilosophy, which was primarily intended to demonstrate the necessity of interdisciplinary cooperation between philosophy and neuroscience. It is, in part, a testament to the success of that earlier work that this new book is her response to the demand for a textbook in neurophilosophy. Unlike many textbooks, however, it it neither impartial nor a survey. Ostensibly, her purpose is to defend three rather weak hypotheses: that mental activity is brain activity, that we need to understand the brain in order to understand the mind, and that neuroscience requires the other cognitive sciences as a guide to what needs explaining. (The first two were far more controversial when she defended them in Neurophilosophy.) The real meat of the book, though, lies in her attempt to show how neuroscience is relevant to a number of traditional philosophical topics, including the self, consciousness, free will, knowledge, representation, and religion. She is rather less successful in meeting this aim than I had hoped, although there is still much to like about the book.

Throughout, Churchland peppers her old enemies with attacks: Fodor's language of thought, dualism, Cartesianism, functionalism, and a priorism of any stripe. It is unfortunate, especially given the textbook format, that her opponents get to speak so little in their own defense. For instance, following a presentation of vector coding in neural networks (Ch. 7), Churchland states that this idea about the nature of mental representation looks much more promising than the language of thought. A naïve reader

might well agree, as one gets not much more than this by way of explanation of the view she opposes: "representations are sentences written in the mind's encyclopedia" (p. 302). The reader is not told in what respects Fodor takes mental representations to be language-like (perhaps he proposes that thoughts are just bits of type knocking about up there?), and so is not in a position to compare the views, or even to tell whether they differ. The key property of language that Fodor attributes to mental representations, namely compositionality (whereby the meanings of complex representations are a function of the meanings of their parts) is not even mentioned. This is especially negligent considering that Fodor has argued in numerous places that if vector coding is compositional, then it's just an implementation of the language of thought, and if it isn't compositional, then it can't serve as a theory of mental representation (Fodor and Pylyshyn, 1988; Fodor, 1998). Perhaps this argument is flawed, but surely it should at least make an appearance.

Some of her other critiques appear to be just bad arguments. Take her treatment of functionalism, for example. This is the view that the identity of a mental state, what makes it the mental state that it is, is the causal role that it plays in the mental and behavioural economy of an organism. Her argument against functionalism (which she equates with the "software view" of the mind) appears to go something like this (see especially pp. 25-27): "Functionalists think that understanding the brain won't help understand the mind, but clearly this is false. So functionalism is false." First, the conclusion is a non sequitur. Functionalism, as such, says nothing about what specific causal roles characterize mental states, or how fine-grained they are - the relevant causal

roles could be at the level of psychophysics, but they could also be at the level of neurophysiology (Lycan, 1987). If the latter, then understanding the brain will most definitely help understanding the mind - but that leaves functionalism, which is a metaphysical thesis about the nature of mental states, entirely untouched. Second, although there have been some foolish souls who deny it (most notably Fodor), most functionalists would be perfectly happy to accept that an investigation of the brain might help reveal the causal roles that characterize particular mental states (see e.g. Prinz, 2000).

Churchland's rejection of functionalism early in the book leads her to seek identities between mental state types and brain state types. For instance, a theory of consciousness, she says, ought to seek an identification of some class of neural activity with awareness (p. 154). She does not even mention the possibility of a functionalist theory of consciousness, despite the fact that such theories have been dominant in contemporary philosophy. While Churchland's comments should persuade us that neuroscience may help show which functionalist theory is true (if functionalism is correct), or which identity theory is true (if that is the correct strategy), she gives us no reason to suppose that neuroscience could help us decide between functionalism and identity theory. That is disappointing, seeing as, in philosophical circles, it is the second thing that is controversial, not the first.

This is just one example where Churchland's strong claims about neuroscience's degree of relevance to philosophy fail to hold up. In the chapter on the self, she moves from the premise that we conceive of the self in many different metaphorical ways (e.g. as

object, as person, as project, and as process), to the conclusion that the self is "a squadron of capacities flying in loose formation", whose nature neuroscience will reveal. From the fact that the ways we conceive of a thing are non-systematic, or even fundamentally confused, it does not follow that there is no coherent, unified thing being conceived. No doubt I have a confused, multifaceted, and metaphorical grasp of spacetime, but if Einstein is right, there is nonetheless a very important, well-defined entity that is spacetime. So while I think she is probably right to recast the problem of what the self is in terms of "self-representational capacities", her motivation for this recasting is very weak. One is better off consulting Dennett, for example (Dennett, 1991).

Similar problems infect Part 2 of the book, on epistemology. (Part 1—which covers the self, consciousness, and free will—is the metaphysics section.) Instead of the nature of knowledge, her target, Churchland tells us how neuroscience will reveal to us the nature of belief (or something like it). But the philosopher also wants to know about justification. Aside from a professed admiration of statistics and decision theory as making real contributions independent of neuroscience (p. 269), Churchland does not even acknowledge that there are legitimate questions about what constitutes justification. Since this is an area in which armchair philosophy is overwhelmingly dominant, her silence on this matter is a serious lacuna in her attack on all things a priori. (At first blush, it certainly seems that neuroscience cannot tell us what it is for a belief to be justified, since we can sensibly ask whether our actual belief forming mechanisms are good ones. This means there must be some standard of justification that is independent of our actual belief

forming mechanisms - though see Kornblith, 2002.)

Also in Part 2, we find an unsatisfying account of representation that ignores the literature (especially Fodor and Lepore, 1999) and, where it is novel, is patently circular. (In describing how a representation's referent is determined, Churchland narrows down the candidate referents by appealing to "what the animal cares about and pays attention to" [p. 307] - but these depend on what the animal represents!) And in discussing whether knowledge is innate, Churchland argues as follows: we are neither blank slates, nor bundles of instincts, because all change (including prenatal development and postnatal learning) causally depends upon both genetic and epigenetic factors. This is hardly germane - nobody thinks that genetic causation is the right way to understand innateness. There are a number of other suggestions on offer (Fodor, 1981; Ariew, 1996; Cowie, 1999; Samuels, 2002), but Churchland does not even acknowledge their existence.

The chapter on free will, in which she argues that a deep neurophysiological understanding of drives, emotion, reward, planning, and decision making will be necessary to understand when an action is voluntary, is more successful. The importance of a deep theoretical understanding is also made clear in the chapter on consciousness (Chapter 4). Her admirable caution here ought to be taken seriously by those engaged in consciousness research, though I think she could go even further than she does. The strategy that currently dominates the scientific study of consciousness, known as "contrastive analysis" (Baars, 1997), is to compare paradigm conscious and non-conscious states, and note their differences. Churchland is not overly sanguine about the potential of

this approach, and rightly so. For one thing, it can confuse background conditions, causes, sequelae, or parallels of conscious awareness for the thing itself (p. 154). More seriously (p. 155):

> We might be looking straight at an instance of the class [of activity identifiable with conscious awareness] without in the slightest recognizing that it is an instance. This will happen if, as is very likely, the physical substrate does not have a property that is salient to the naive observer, but is recognizable only through the lens of a more comprehensive theory of brain function.

She uses an analogy (of which I am also fond) to illustrate this possibility: only through the lens of a comprehensive theory does the identity of light and electromagnetic radiation become apparent. I think we can draw a direct analogy with so-called "contrastive analysis" studies, which always rely on a single index of consciousness, namely introspective report. It could be that introspective report is systematically misleading, just as our vision misses a huge swath of the phenomenon that is light.

Churchland's down-home writing style is engaging and pleasant to read, which will no doubt be appreciated by students. While the level of argumentation in the book is inadequate for a course in which argumentation is the focus (i.e. most philosophy courses), her descriptions of the relevant neuroscience are interesting and clear. (Some small exceptions are the rather opaque discussion of spatial representation in Chapter 7, and the second half of chapter 8, which reads like a random collection of facts about the hippocampus, with an incredibly brief and uniformative discussion of learning algorithms

tacked on at the end.) You will have gathered, though, that I think Churchland has limited success in achieving her main goal, namely to show how neuroscience can provide or at least point towards solutions to age-old philosophical questions. (This despite the fact that I agree with many of her conclusions.) No doubt I am wrong about some of the flaws in her arguments (and there are more than I have discussed here), but if so, it is at least in part because the arguments are so well hidden in the text - extracting them is more of a chore than it should be.

That said, there is still much to like about the book. She is at her best, not when trying to contribute to what she seems to regard as rather sterile and pointless debates in philosophy, but when showing how psychological features might be explained at the neural level (e.g. body representation, motor control, binocular rivalry, emotion, appetite, and face recognition). While there is little that is original in these explanations (not surprising in a textbook, and not her intention), a researcher wanting to learn more about how psychology and neuroscience relate could do worse than beginning with this book.

Ariew, A. (1996). Innateness and Canalization. *Philosophy of Science*, *63(Proceedings)*, 519-29.

Baars, B. J. (1997). *In the theater of consciousness : the workspace of the mind*. New York: Oxford University Press.

Cowie, F. (1999). *What's Within? Nativism Reconsidered*. Oxford: Oxford University Press.

Dennett, D. C. (1991). *Consciousness explained*. Boston: Little, Brown and Co.

Fodor, J. A., & Lepore, E. (1999). All at Sea in Semantic Space: Paul Churchland on meaning similarity. *Journal of Philosophy*, *96*, 381-403.

Fodor, J. (1981). The Present Status of the Innateness Controversy, in *RePresentations: Philosophical essays on the foundations of cognitive science* (pp. 257-316). Cambridge, Mass.: MIT Press.

Fodor, J. (1998). *Concepts: where cognitive science went wrong*. Oxford: Oxford University Press.

Fodor, J., & Pylyshyn, Z. W. (1988). Connectionism and Cognitive Architecture: A Critical Analysis. *Cognition*, *28*, 3-71.

Kornblith, H. (2002). *Knowledge and its place in nature*. Oxford: Clarendon Press.

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

Prinz, J. (2000). A neurofunctional theory of visual consciousness. *Consciousness and Cognition*, *9(2)*, 243-59.

Samuels, R. (2002). Nativism in Cognitive Science. *Mind and Language*, *17(3)*, 233-65.