Jesse Prinz, *The Conscious Brain: How Attention Engenders Experience*. New York: Oxford University Press, 2012. xiii +397 pp.

In his superb book *The Conscious Brain*, Prinz defends a theory of consciousness: a theory that (among other things) tells us what kind of thing consciousness is, what distinguishes different kinds of conscious states, and how consciousness is implemented in the human brain. Prinz's view is that conscious experiences are attended intermediate-level perceptual representations. Intermediate-level representations are, roughly, the kinds of representations that Marr (1982) had in mind when he discussed the "2 ½ D sketch": a 3D world is represented, but not in a way that abstracts from the viewpoint of the observer. For Prinz, such a representation is attended just if the information it represents is accessible to one of the set of postperceptual systems known as "working memory." The relevant representations are "vectorwaves" (coordinated patterns of neural firing that code information about features of the environment), and their attentive availability is implemented by synchronized firing in the gamma range across the relevant neural populations. This gives us an "unpacked" version of the view that provides an unusually vivid sense of how consciousness might actually be implemented in our neuronal wetware. The resulting discussion, drawing on a vast range of empirical data and astute philosophical reasoning, is absolutely essential reading for anyone interested in the topic. The comprehensiveness of topics discussed and the lucid and simple prose also make the book of great pedagogical value.

Prinz's view contrasts with views on which different *kinds* or different *degrees* of accessibility to postperceptual cognition are required for information to be conscious, or accessibility is not required at all. For example, Block's (2007) view is that consciousness doesn't constitutively involve postperceptual cognitive mechanisms at all; he thereby allows for conscious experiences that are more inaccessible to these systems than Prinz's view allows for. On the other extreme, some views require that conscious perceptual information be actually encoded in these postperceptual systems, not merely accessible to them; for example, this is what we get on Dehaene and Naccache's (2001) neuronal global workspace model. The view also differs from views on which conscious experience doesn't exclusively involve intermediate-level content. For example, it might also involve more high-level representations, such as those involved in object categorization, as on Siegel's (2010) view.

Both aspects of the view raise epistemic and explanatory problems. What is the right methodology for figuring out when consciousness arises in the brain? Prinz's methodology is to figure out empirically the conditions under which states in the brain fit certain pretheoretical functional criteria we associ-

Thanks to Ned Block for helpful comments on an earlier draft.

ate with consciousness (in particular, forms of access-consciousness like a subject's ability to report information), and then conclude that these are the conditions under which consciousness exists. This makes it unsurprising that for him, consciousness is a kind of accessibility. One problem with this is that arguably there is no a priori guarantee that there aren't conscious states that don't fit the pretheoretical criteria (as Chalmers [1998] pointed out). How then do we know that there aren't conscious states that are not accessible to postperceptual cognition? A second problem is that even if the method correctly tells us when consciousness arises in the human brain, how do we know it isn't merely uncovering contingent features of how consciousness is *implemented* in us, rather than *essential* features? We want to know what consciousness *is*, not merely when it contingently occurs in us. A third kind of worry is explanatory. It is tempting to ask: *why* does consciousness have the features Prinz says it does? This becomes especially salient when we see that some of these features seem to be quite arbitrary, as I will explain.

Why does Prinz reject inaccessible, unreportable experiences? Ultimately, his reasoning turns on something like the following principle: unless we have positive evidence that such inaccessible states exist, then we are justified in believing that they *don't* exist (105, 107–9). But although this is not implausible, it is unclear what exactly justifies this principle. It might seem like a straightforward application of theoretical parsimony. But on reflection, given the existence of some representation in the brain, saying that it's conscious isn't any less parsimonious that saying it isn't. Inaccessible experiences aren't like invisible ghosts; perhaps a better analogy is discovering a new kind of particle and not knowing whether it is positively charged. Second, the method of pretheoretically associating conscious experience with forms of access-consciousness will likely prevent us from finding evidence for inaccessible states, so an absence of evidence is hardly evidence of absence in this case. Finally, Prinz's view of phenomenal concepts seems to commit him to agreeing with Chalmers that there is no conceptual or semantic guarantee that there aren't conscious states that fail to fit our pretheoretical functional criteria. So we are left wondering what justifies Prinz's basic epistemic principle.

The kind of accessibility that Prinz associates with consciousness also invites explanatory puzzlement. For example, you might think that being accessible to working memory would require that working memory can actually use the relevant intermediate-level information. But, on the contrary, it turns out that working memory doesn't use intermediate level information *at all*, but rather more abstract high-level information. The intermediate-level information is "accessible" only in the sense that it is the input into a process of high-level categorization that leads to high-level representation in working memory (101–2). This makes one think—what's so special about this rather awkwardly indirect form of accessibility? Why is *that* what constitutes consciousness?

Prinz could reply that consciousness occurs at the intermediate level and that this is the kind of accessibility that applies to these representations. But, of course, that invites the reply: what's so special about the intermediate level? Although Prinz argues that, *in fact*, all conscious representations are intermediate level, he doesn't explain *why* consciousness is constrained in this way. One thing that makes this especially puzzling is that he rejects representationalism, holding that qualitative character is determined by intrinsic features of the representational vehicle (a "vectorwave"), not its content (11–21, 126–33). But if informational content is irrelevant to qualitative character, why does it matter for consciousness?

One can ask the same kind of explanatory questions about other features of Prinz's view. How might he respond? Rather than tackling them head on, he might reject such questions as illegitimate or based on false presuppositions. For example, he might say that theoretical identities are not the kind of thing that can be explained—more on this below. Another option is to say that he is only telling us how consciousness is contingently *implemented*, not what it is essentially. For example, maybe there's nothing about consciousness that requires it to involve only intermediate-level representations; it's just a contingent fact that *in us*, these are the only representations that meet the relevant conditions (Mole [2013] suggests this retreat in his review). Retreating to a view of implementation rather than essence gives the account less explanatory depth, but at least it might avoid some of these hard explanatory questions. It raises the question: how do we draw the line between essence and implementation?

Prinz (2003) has previously taken a mysterian approach to this issue, holding that this dividing line is unknowable; one could also be deflationary about it (see below). But in *The Conscious Brain*, he thinks we can do better. The relevant level is "a level of analysis that can account for every aspect of conscious experience, in the sense that every conscious difference can be mapped onto a difference at that level" (289). The idea is supposed to be that absent positive evidence that a physical feature is essential, we should assume it isn't (echoing the methodology mentioned above); in effect, we should take consciousness to be as high level as we can. This suggests that Prinz might indeed err on the side of saying that the features of his account concern only how consciousness is contingently implemented.

However, we can question Prinz's principle here. We don't seem to adopt it in other cases; for example, taking water to be H_20 probably wasn't taking the most high-level view available. Furthermore, what counts as a "conscious difference"? If we're talking about differences in what someone reports, or discrimination behavior more generally (and this appears to be what he has in mind), then what differences are relevant? Even very fine-grained features of neural states carry information about the environment and can have a small statistical impact on discriminatory abilities. So if we want to explain *every* aspect of a

subject's discriminatory capacities, we'll need to look at the brain at a very fine-grained level. But if only less fine-grained statistical features of their discriminatory behavior are relevant, then what's the relevant level of grain? On the other hand, if we're talking *directly* about experiential differences, rather than discrimination behavior, then the problem is that we don't have any direct way of measuring these differences, so we are pushed back toward his earlier mysterian view.

Prinz could also respond to the question "why does consciousness have feature X?" by saying that theoretical identities *can't* be explained. Consciousness just is what it is. In my view, this doesn't satisfactorily explain away the sense of unacceptable arbitrariness in aspects of the view (for example, why only intermediate-level representations?). What else might he say? Prinz expresses an interestingly deflationary attitude toward theoretical identification in other cases. He writes, "Every theoretical identification involves some degree of arbitrary decision" and that "the vast majority of theoretical identities by scientists are actually pragmatically and politically brokered decisions rather than discoveries" (299). If he took this view with consciousness, he could say that details of his view appear arbitrary and explanation demanding because they really *are* arbitrary, in the way that *all* theoretical identities are.

But his view is that the case of consciousness is quite different from this: "We assume that scientists are not making an arbitrary choice when they decide between different neural correlates" (300). I suspect that Prinz should take more seriously a similarly deflationary attitude toward questions about the essence of consciousness. Imagine a Martian scientist explaining our behavior by reverse engineering the human brain. It's not as if one of the consciousness-candidates under discussion here is going to stand out to them as having a special glow. Relatedly, if they are trying to interpret our term "consciousness," it might seem rather arbitrary and uninteresting what *exactly* the interpretation of the term ought to be. If we are materialists and we think that all that really exists are the kinds of facts that this Martian scientist would know about, shouldn't we have a fairly deflationary attitude toward finding the nature of consciousness that mirrors how things would look to the Martian scientist?

Compare the case of life: we might be worried about how life could possibly be a physical phenomenon. But once we have overcome these worries, it seems to be a mistake to worry about exactly what the necessary and sufficient conditions for it are. We learn what life is by learning about the physical processes that underwrite its specific manifestations around us. Who cares how much of this is strictly necessary for "life"? (Notice that even if we are deflationary about what exactly life is, that doesn't require us to think that it is unimportant: for example, we could still think that it is astonishing that the universe contains such a thing.) Similarly, we learn a lot about what consciousness is by learning about the specific processes that underwrite it in the brain, but perhaps there's no deep fact about which parts of this processing are really required (this

is not to say that it's *completely* indeterminate: some views might be determinately wrong, just as some views of what life is are determinately wrong). Adopting such a perspective might enable Prinz to offer deflationary responses to both the epistemic and explanatory challenges I raised earlier.

Against this, it's true that Prinz argues that the analogy between reducing consciousness and other theoretical reductions breaks down because of the unique kind of epistemic access we have to consciousness. The issues here are complex, but I suspect it is a mistake to think that this *epistemic* asymmetry justifies a more inflationary *metaphysical* view. The real significance of the epistemic disanalogy comes in explaining, or explaining away, the philosophical puzzlement surrounding consciousness: how *could* it be something physical? Even if we are deflationary about the metaphysics of consciousness, there is still the highly significant philosophical task of ridding ourselves of this explanatory puzzlement.

Despite these misgivings, I still think this is a great book. If you haven't read it yet, you should. If you have already read it, read it again!

References

Block, N. 2007. "Consciousness, Accessibility, and the Mesh between Psychology and Neuroscience." *Behavioral and Brain Sciences* 30, no. 5–6: 481–99.

Chalmers, D. J. 1998. "On the Search for the Neural Correlate of Consciousness." In *Toward a Science of Consciousness II: The Second Tucson Discussions and Debates*, ed. S. Hameroff, A. Kasniak, and A. Scott, 219–29. Cambridge, MA: MIT Press.

Dehaene, S., and L. Naccache. 2001. "Towards a Cognitive Neuroscience of Consciousness: Basic Evidence and a Workspace Framework." *Cognition* 79, no. 1: 1–37.

Marr, D. 1982. *Vision: A Computational Approach*. New York: W. H. Freeman.

Mole, C. 2013. "Review of Prinz, *The Conscious Brain*." *Notre Dame Philosophical Reviews*. http://ndpr.nd.edu/news/36606-the-conscious-brain-how-attention-engenders-experience/.

Prinz, J. 2003. "Level-Headed Mysterianism and Artificial Experience." *Journal of Consciousness Studies* 10, no. 4–5: 111–32.

Siegel, S. 2010. *The Contents of Visual Experience*. Oxford: Oxford University Press.

Geoffrey Lee
University of California, Berkeley

Philosophical Review, Vol. 124, No. 1, 2015
DOI 10.1215/00318108-2812711