

Physicalism, or Something Near Enough, by Jaegwon Kim. Princeton: Princeton University Press, 2005. Pp. xiii + 186. H/b £17.50.

In this compact and readable book, Jaegwon Kim provides an overview of his considered position on the mind-body problem, updating and refining several familiar arguments as well as introducing newer material to round out his map of the terrain. The position 'near enough' to physicalism endorsed at the end holds that, with the exception of the intrinsic features of phenomenal states, mental properties are physically reducible via functional analysis. The unreduced features are epiphenomenal, Kim stresses, but we can live with this result if we bear in mind that intentional properties and the relational features of qualia can still play a causal role.

A brief part of the final chapter is devoted to sketching the positive quasi-physicalist view; the rest of the book may be seen as defending two main theses. The first thesis is made explicit by Kim as 'conditional reductionism' (p. 161). This is the claim that mental properties must be epiphenomenal if not physically reducible. The other thesis is that a defense of reduction requires a kind of functional analysis of the properties to be reduced. While Kim does not make it explicit, it seems clear that the sort of functional analysis he has in mind must be a priori in character. If he is right that such analyses are out of the question for phenomenal properties but available for other mental properties, then the two theses will land us in Kim's quasi-physicalist position.

The defense of conditional reductionism consists of Kim's well-known 'supervenience argument' as well as an attack on substance dualism. The supervenience argument aims to show that non-reductive materialists cannot accommodate mental causation; the attack on substance dualism in effect follows up on this by trying to show non-reductive *non*-materialists face the same difficulty. As a result, only a reductive materialist can avoid epiphenomenalism.

Kim's defense of the claim that reduction requires a kind of functional analysis occupies much of the second half of the book, wherein Kim argues that a posteriori necessary identities of the sort made famous by Kripke are not in fact suitable for the sort of reductive explanation we need. Only a functional analysis will do as a response to the explanatory gap that originally motivates the dualist project. Kim's primary argument turns on the claim that such identities cannot play the explanatory role they are alleged to play in recent work on accommodating phenomenal properties in a physicalist framework.

These arguments will undoubtedly stimulate a great deal of debate. The material is presented in a brisk and bold fashion, and Kim's despair of finding an adequate physicalist treatment of qualia will surely be found provocative. In my comments I focus on three things: the attack on substance dualism; the operative notions of reduction and reductive explanation; and the argument for thinking that a posteriori identities cannot do the work needed to save reductionism.

Kim claims that the substance dualist cannot make sense of causal interaction between immaterial minds and physical bodies—or even between one

immaterial mind and another. The fundamental problem stems from the lack of spatial location of such immaterial minds. As Kim sees it, there must be an account of the 'principles that underlie the correct and incorrect *pairings* of cause and effect' (p. 79), and such a 'pairing problem' cannot be solved without the availability of either spatial relations or some analogous family of relations. Since mental substances are not located in space, no relations of contiguity or the like are available to ground causal relations; nor do there seem to be any analogous relations that can do the job.

The argument is not compelling. A simple reply is to take causal relations as primitive; no principles are needed to 'underlie' these pairings of cause and effect. The 'pairing relation' that holds between every pair of cause and effect is just causation itself. Strangely, Kim does not take note of this possibility. Perhaps he takes it as obvious that causation itself must be reducible to some other relations, but this surely should not be taken for granted. Further, even if we decide that spatial relations are needed, it remains unclear to me why a substance dualist should not go ahead and locate such substances in space. Kim briefly considers this move; his main criticism seems to be that any location we posit would be arbitrary and ad hoc. At one point he writes:

It would beg the question to locate my soul where my body, or brain, is on the ground that my soul and my body are in direct causal interaction with each other; the reason is that the possibility of such interaction is what is at issue and we are considering the localizability of souls in order to make mind-body causation possible. (p. 89)

Pace Kim, this is not begging the question at all. It may be that what comes first epistemically is the knowledge that two entities are causally related, whereas what grounds the relation is a spatial relation we only later infer.

Let us move on to a more central matter: the notion of reduction. At the start of chapter four Kim distinguishes between reduction and reductive explanation. While no formal account of the difference is given, the spirit of the distinction seems clear. Whereas reduction requires the truth of some ontological claim—such as an identity claim of some sort—a reductive explanation is an explanation of some phenomenon in terms of phenomena at a 'lower level', where that explanation may or may not require an accompanying ontological claim. The model is, roughly, one whereby a set of statements describing the explanatory (lower level) phenomenon can be used as premises to derive a statement describing the phenomenon to be explained. Kim introduces the distinction as providing two ways one might respond to the infamous explanatory gap. He does not, however, make plain how the distinction relates to the general thesis of conditional reductionism. Should we understand that thesis as the claim that mental properties, if they are not to be epiphenomenal, must be reduced to the physical? Or as the claim that to avoid epiphenomenalism they must be subject to an appropriate sort of reductive explanation?

The logic of the supervenience argument suggests saving mental causation requires reduction, not reductive explanation. In that argument, the premise

of irreducibility manifests itself simply as the denial of any type identity between mental and physical properties (p. 42). On the face of it, then, reduction of a mental property *M* amounts to nothing more and nothing less than the positing of an identity between *M* and some physical property.

Matters are more complicated than this, however, as is made clear by Kim's treatment of multiple realizability. That treatment comes up here in the context of responding to the worry that the supervenience argument generalizes to biological, chemical, and other non-mental, non-physical properties. He stresses that the argument is indeed meant to generalize to any property not subject to physical reduction, though he does not find it plausible that biological and other non-mental properties are thus irreducible. One might suspect them of irreducibility because of their multiple realizability; on Kim's view, however, 'multiple realizability only leads to reducibility to multiple reduction bases, not to irreducibility' (p. 56). One straightforward way to understand this move is as a kind of eliminativism. Let *M* be the multiply realized property; the idea is that *M* is eliminated in favor of several surrogates, each of which is only singly realizable. Thus, we have *M-for-structure-S₁*, *M-for-structure-S₂*, and so on, where each of these is identical with some physical property, namely, that which uniquely plays the right role in that sort of structure. If this is Kim's view, however, then we apparently are not preserving the causal efficacy of the original *M* but only that of its various surrogates.

It is not clear to me, however, that this is quite what Kim has in mind in his current presentation of the supervenience argument, as a few pages later he writes:

All we need is identity at the level of instances, not necessarily at the level of kinds and properties; causation after all is a relation between property or kind-instances, not between properties or kinds as such. (p. 58)

The claim that each mental property *instance* is identical with some physical property instance seems not to imply any type identity thesis at all. If Kim's 'conditional reductionism' requires only this much by way of reduction, it is not clear that those who call themselves non-reductive materialists are in fact opposed to such reduction. What is plain is that the non-reductive materialist wants to insist that (i) there are indeed mental properties and (ii) they are not identical with physical properties. These claims are *prima facie* consistent with identity 'at the level of instances'.

However exactly we understand what Kim means by 'reduction,' we face a puzzle regarding the role of reductive explanation in his overall argument. It seems plain that one way, at least, of being a reductionist is to maintain a type identity thesis. Suppose that, in response to conditional reductionism, we opt for such a thesis. What then is the role of Kim's remarks on reductive explanation? It may seem at first that Kim argues as follows. First, establish conditional reductionism; second, infer that to save a mental property from epiphenomenalism, it must be subject to an appropriate reductive explanation; third, argue that while many mental properties are subject to such explanation, phenome-

nal ones are not, and hence must be deemed epiphenomenal. But this cannot be Kim's argument, for he stresses that type identities *remove* the need for any reductive explanation. In his discussion of Block and Stalnaker, he imagines them positing such identities and then saying this:

Identities like [the claim that consciousness is identical with pyramidal cell activity] should be seen not as helping to answer explanatory questions like 'Why is Jones conscious whenever pyramidal cell activity is going on in his brain?' but rather as neutralizing or dissipating them—that is, as showing that *there is nothing here to be explained*. (pp. 116–17)

A page later Kim describes his 'main point' thus:

On Block and Stalnaker's account of identity-mediated reduction, even if we have successfully reduced pain to C-fiber stimulation ... that would not yield an explanation, reductive or otherwise, of the occurrence of pain in terms of C-fiber stimulations. (p. 118)

In view of this, how exactly does Kim motivate the claim that avoiding epiphenomenalism requires a functional reductive explanation? His view, I suspect, is not that conditional reductionism *requires* of causally efficacious properties that they be subject to reductive explanation; it is rather that, first, such properties must *either* be identical with physical properties *or* be subject to an appropriate reductive explanation, and second, in the case of mental properties, the former option is independently implausible. As a result, the only plausible option for the would-be reductionist about mental properties is to offer a reductive explanation. The final step in the argument is to maintain that such explanations are available for intentional properties, and even the relational features of phenomenal states, but not the intrinsic phenomenal properties themselves.

If we understand Kim this way, the question arises: how is the conditional reductionist thesis, understood in this specifically disjunctive way, to be justified? The main argument Kim gives for this thesis—the supervenience argument—seems to require identity, not identity *or* reductive explanation. I offer one speculative remark: perhaps Kim would like to say that causal efficacy requires either type identity (which disallows reductive explanation) or identity 'at the level of instances'—which, perhaps, allows us to make sense of reductive explanation. While it may be senseless to ask why property *M* is present whenever property *P* is present when *M* is identical with *P*, perhaps it makes sense to ask why a particular instance of *P* is also an instance of *M*, given that the *types M* and *P* are distinct.

Be that as it may, let us turn to one final matter: Kim's critique of explanatory arguments for type identity. One important preliminary point is in order. Kim sets up an opposition between functional reduction and type identity; he seems to associate this contrast with a further, distinct one, namely, the contrast between a priori and a posteriori accounts of the mental. The association is not mandatory. One could defend an identification of a mental property with a *functional* property on a posteriori grounds just as one could an identification with some physical or neurological property. Kim's fifth chapter

(‘Explanatory Arguments for Type Physicalism and Why They Don’t Work’) should, I think, be understood as attacking a style of argument aimed at either sort of a posteriori identity thesis.

The style of argument in question can be characterized as having the following form. If we posit such identities, we can produce explanations we could not produce before. The identities are then justified by being key parts of explanations that are themselves justified by inference to the best explanation. According to Kim, however, these identities do not in fact play any key role in those explanations; they serve merely as ‘rewrite rules’ that allow us to re-describe the original explanatory facts. His point is illustrated by the following example:

Tully is wise.

Tully = Cicero.

Therefore, Cicero is wise. (p. 132)

Kim’s comment is that no one would take this seriously as an explanation of why Cicero is wise. It is not implausible to say that the fact that Tully is wise is the very same fact as that Cicero is wise, in which case it is plain that no explanation of *that fact* has been offered. Similarly, we should not suppose we can explain, say, the fact that Jones is in pain at *t* by adverting to (i) the fact that Jones’s C-fibers are firing at *t* and (ii) the fact that pain is identical with the firing of C-fibers. Positing the identity of pain with C-fiber firing allows us to see that there is only one fact here, but it does not explain anything.

Whether Kim is right about this depends on how we individuate facts. If indeed ‘Tully is wise’ and ‘Cicero is wise’ describe the same fact, then presumably ‘Tully is identical with Cicero’ and ‘Tully is identical with Tully’ describe the same fact as well; and it is surely no surprise that the fact that something is identical with itself is incapable of playing any explanatory role.

Supposing Kim is right about this, however, it seems to me that there is another, quite natural way to construe these explanatory arguments that does not run afoul of Kim’s critique. The assertion of ‘Tully is identical with Cicero’ will convey something more than just the fact that Tully is identical with himself; it will also convey the fact that the names ‘Tully’ and ‘Cicero’ refer to the same thing. This sort of meta-linguistic fact may well play the requisite explanatory role. Consider McLaughlin’s version of the argument, where the *explanandum* is the fact that a given mental property is correlated with a given physical property, for example the fact that in general someone is in pain if and only if his or her C-fibers are firing. The identity of pain and C-fiber firing is then offered as a way to explain this. If Kim is right, however, then identity cannot do that sort of explanatory work. But now consider by contrast the alleged fact that ‘pain’ and ‘C-fiber firing’ co-refer. Could that fact explain the fact that someone is in pain if and only if his or her C-fibers are firing? No. But it could explain why statements of the form ‘*x* is in pain’ and statements of the form ‘*x*’s C-fibers are firing’ are true of the very same individuals. And that,

I suspect, is enough to restore the explanatory arguments for a posteriori identifications.

Department of Philosophy
University of Florida
P. O. Box 118545
Gainesville, FL 32611-8545
Florida
USA
doi:10.1093/mind/fzl136

D. GENE WITMER

The Knowability Paradox, by Jonathan Kvanvig. Oxford: Oxford University Press, 2006. Pp. 240. H/b £65.00.

This is, to my knowledge, the first book-length treatment of a short proof, first published by F. B. Fitch in 1963, which is now commonly known as the ‘paradox of knowability’. But this book should not be taken to have a narrow target audience just because it is all about a single proof. The issues discussed are significant for many areas of metaphysics and the philosophy of logic and language. Anyone thinking about the nature of truth, modal logics, realism and anti-realism, the semantics of quantification or our capacity to know about the world will find something of interest here.

In addition to proposing a ‘neo-Russelian’ treatment of quantification in an attempt to resolve a puzzle which he thinks Fitch’s proof raises (ch. 6), in his first two chapters Kvanvig offers an extended discussion of the proof and its significance. He then reviews some of the extant literature on the topic in chapters three, four and five. Kvanvig’s style, forthright and at times provocative, is engaging whether or not one is sympathetic to the claims being argued for, and the structure of the book is continuous and well signposted. Although I have some reservations about the book’s major claims, these should not distract from the interest and importance of both the review of extant literature and Kvanvig’s own proposal.

One major claim of the book is that the real import of Fitch’s proof has not hitherto been properly understood. The proof begins with the premiss that all truths are knowable, which is often taken to be a commitment of global anti-realism, and proceeds to the conclusion that all truths are known (by someone at some time). It does this using only apparently harmless logical steps, and appealing only to two intuitive principles about knowledge: factivity and distributivity. The conclusion looks obviously false. The proof, therefore, is usually viewed primarily as a threat to anti-realists, who are taken to be committed to its premiss.

But Kvanvig argues that, regardless of its significance for anti-realism, a deeper puzzle is raised by the proof, and it is a puzzle for everyone. In his opinion (p. 2), the proof shows that: