Mental Causation and Intelligibility

David Robb[†]
darobb@davidson.edu

ABSTRACT

I look at some central positions in the mental causation debate – reductionism, emergentism, and nonreductive physicalism – on the hypothesis that mental causation is intelligible. On this hypothesis, mental causes and their effects are internally related so that they intelligibly "fit", analogous to the way puzzle pieces interlock, or shades of red fall into order within a color sphere. The assumption of intelligibility has what I take to be a welcome consequence: deciding among rivals in the mental causation debate could end up to be largely an empirical matter.

Keywords: reductionism, emergentism, nonreductive physicalism, mental causation, causal powers

1. Introduction

In this paper I look at some central positions in the mental causation debate – reductionism, emergentism, and nonreductive physicalism – on the hypothesis that mental causation is intelligible. While intelligibility on the face of it is an epistemic notion, it has a metaphysical side as well: intelligible causation is an internal relation among the causal relata, here construed as powers. These internal relations make possible the epistemic side of intelligibility: grasping the nature of causes and their effects, we can see how the former produce the latter, in the same way that grasping the nature of, say, colors, we can see how they fall into place within a color sphere along the dimensions of hue, saturation, and brightness. If mental causation is intelligible, mental causes are likewise internally related to their behavioral effects, making possible similar epistemic feats.

In the sections to follow, I will not try to show that mental causation, or indeed any kind of causation, is intelligible. Instead, my more modest aim is to

[†] Davidson College.

argue that the working assumption of intelligibility could lead to an empirical resolution of the mental causation debate. I consider this a welcome result, as standard philosophical lines of argument, having advanced the debate up to a point, are less likely to make further headway on these issues. That said, the discussion below is not philosophically neutral. As noted already, I take for granted a powers conception of causality, speaking freely of mental and physical properties and the powers they bestow.[1] Moreover, I assume that there is mental causation, that mental causes enter into the production of intentional behavior: epiphenomenalism and other forms of non-causalism will not play a part below. These background assumptions, as controversial as they are, leave open a number of options in the metaphysics of mental causation, and this is what I explore in what follows. I hope to show that the hypothesis of causal intelligibility makes empirical evidence more relevant to deciding among these options than may initially be supposed.

2. Intelligibility

Start with some standard examples of intelligible causation.[2] Two gears of the same radius are meshed; as one turns clockwise, it causes the other to turn counterclockwise at the same rate. A square paperweight is pressed into soft clay, thereby producing a square impression. A sphere on an inclined plane rolls smoothly, unlike a dodecahedron, which clunks down the same plane. In cases such as these, it seems as if cause and effect intelligibly "fit", analogous to the way puzzle pieces interlock, or colors fall into order within a color sphere.

So-described, intelligibility appears to be epistemic notion, and so it is. When one sees how puzzle pieces fit, or how colors are arranged in a color sphere, this "seeing" is epistemic. But here one is tracking a metaphysical fact about relations among properties. The shapes of the pieces, and the colors themselves, are internally related so that necessarily, any pieces with those shapes, or colors of those natures, will fall into the same arrangements. In

[1] For recent defenses of the powers conception of causality, see Mumford and Anjum (2011); and Heil (2012, ch. 6). In what follows, it will sometimes be useful to speak of a property, sometimes of the powers it bestows. However, under the assumption of causal intelligibility, a property and its powers – i.e., its potential to combine with other properties to produce effects – are one and the same. For more on this "identity" theory of powers, see Heil (2003, ch. 11), and for an application of the identity theory to mental causation, Robb (forthcoming).

[2] Primary inspiration comes from Sanford (1975; 1994); other examples and references are in Robb (2007).

causation, the internally related properties are *powers*. When one gear turns another, its properties, such a shape and rigidity, are intelligibly manifested. These powers are internally related so that any objects with those powers must produce the effects as indicated – or so claims the thesis of causal intelligibility.

Many philosophers, perhaps under Hume's influence, will balk at this thesis. Humean worries about intelligibility must be confronted, but this is not the place for it, as I aim, not to defend causal intelligibility, but rather to explore its consequences for the mental causation debate. That said, so as not to make this hypothesis more controversial than it needs to be, I should point out, first, that there's nothing in the thesis of intelligibility requiring that causal relations can be known *a priori*. Sometimes a structure of internally related properties can be revealed only through empirical investigation. The causal structure resulting from the nature of powers could turn out to be like this; it need not be revealed to us through a priori reflection on our concepts. Second, in what follows I do not require a global thesis of causal intelligibility. All that's needed is the working assumption that the mental causation of behavior is intelligible. It's compatible with this that other sorts of causation – such as causation at the most basic physical level – is not intelligible. For example, perhaps it's not intelligible that like charges repel. But it could turn out that given such microphysical facts, the mechanisms of mental causation are intelligible. By analogy, the pixels making up the pieces of a jigsaw puzzle might not be intelligibly related. It could appear brute that pixels of a puzzle are adjacent. But given the way pixels are arranged to form pieces, it is intelligible how one piece fits with another. It's this sort of "macro-intelligibility" that concerns me in this paper, whether or not relations among powers are intelligibly related at every mereological level.

The bulk what follows is structured around two alternative scenarios, each a potential outcome of a sufficiently advanced science of intentional behavior. (I will not weigh in on which scenario has or will come to pass.) I'll argue that given the assumption of causal intelligibility, these scenarios, were they to be realized, could empirically distinguish among the main rivals in the mental causation debate.

3. First Scenario: The Causation of Behavior is Physically Intelligible

In the first scenario, mental causation is found to be *physically* intelligible. That is, behavior is discovered to be the intelligible manifestation of powers, all of which are known to be physical. To our future scientists, behavior "fits"

with its physical – presumably biochemical – causes as plainly as the pieces of a jigsaw puzzle. Here, apparently, is empirical confirmation of reductionism, the thesis that mental powers are wholly physical. Mental causation is revealed as merely a complex form of macro-physical causation. There are no "top-down", non-physical causes operating in this first scenario, for if there were, mental causation would not appear physically intelligible. Finding no such gaps in the causal history of behavior, our future physiologists can reasonably conclude on empirical grounds that mental causation is just a kind of physical causation.[3]

One may object that this first scenario could never come about, that the causation of behavior could not be physically intelligible, for behavioral properties are higher-level properties at best realized or constituted by physical properties. Supporting such a claim are the standard ways of distinguishing behavior from bodily movement, including the multiple realizability of the behavioral in the physical. If intelligible causation occurs only within a single level, we could never discover, says the objection, that the mental causation of behavior is physically intelligible. Now I'm skeptical of the arguments placing behavior at a higher, more abstract level. They frequently seem to turn on an illegitimate slide from semantic or epistemic premises to metaphysical conclusions.[4] But I needn't here enter into this controversy. Even if the physical effects of mental causation merely constitute behavior, this will not preclude our first scenario so long as constitution is itself intelligible, as it certainly appears to be. Behavioral properties, that is, are internally related to the physical features that constitute them. It's plausible in fact that behavioral properties logically supervene on their physical properties – including, if necessary, "wide" physical properties – and logical supervenience is a paradigmatic intelligible relation.[5] Any reductionist arguments from the physical intelligibility of mental causation should go through just as well whether the effects are strictly behavioral or merely behavior-constituting physical processes.

Empirically resolving the mental causation debate – for the moment, in favor of reductionism – looks promising. However, matters cannot be as simple as presented so far, for this first scenario may not be as friendly to reduction-

[3] This has affinities with the causal argument for reductionism advanced by Papineau (2001) among others. The causal argument turns on the closed character of the physical world (or at least that portion studied by the science of physiology). I am sympathetic to the causal argument, but for reasons given at the end of this section, I believe an empirical case for reduction is best framed in terms of physical intelligibility rather than causal closure.

[4] See Heil (2003, ch. 3).

[5] For a general discussion of logical supervenience, see Chalmers (1996, ch. 2).

ism as it initially appears. There are in fact at least two non-reductionist positions that look compatible with the envisioned discovery. First is non-reductive physicalism (NRP), on which the causal powers of mental properties are included among the powers of their physical base properties. Second is a version of emergentism on which the novel, top-down powers bestowed by non-physical properties work smoothly within the physical world so as to present the appearance of reduction. I will consider such views in turn, arguing that the working assumption of intelligibility, albeit with some needed supplements, could let us rule out even these rivals to reductionism.

Begin with NRP which, as I'll construe it, says that all mental properties are realized in physical properties. Here realization is a relation intimate enough so that mental properties are necessitated by their physical base properties, but weak enough to preclude identity, and thus reduction.[6] If NRP is not to collapse into emergentism, the causal powers of mental properties should not be novel, but should instead be included among the causal powers of their realizing physical properties. (Such inclusion is a necessary component of any view deserving the physicalist label.) But now it looks as if NRP fits at least as well as reductionism with the physical intelligibility of mental causation. After all, on both views, the same set of physical powers produce behavioral effects. The only difference here is that on NRP, these powers are also bestowed by irreducible mental properties; but there's no reason at the outset to think this addition of matching powers would disrupt the physical intelligibility of mental causation. The upshot is that the mental causation debate is not likely to be resolved under the first scenario after all: both reductionism and NRP remain live options.

So far I've described NRP as if it presented a single picture of mental powers, but in fact NRP's proponents have at least two options that should be distinguished. Consider a mental property *M* and the physical property *P* realizing it on a given occasion. *M*'s powers could be "included" among *P*'s powers in either of two senses, each resulting in a version of non-reductive physicalism. On NRP-1, each of *M*'s (token) powers is identical to one of *P*'s powers. Here the powers of the mental property are wholly contained within those of its physical base. Alternatively, on NRP-2, *M*'s (token) powers are distinct from any of *P*'s powers. The powers of the mental property are, in that sense, non-physical. But each mental power's causal profile – that power's potential to combine with others in the production of behavior – matches the causal profile of some power

[6] Non-reductive physicalism has many contemporary adherents; classic defenses include Putnam (1980) and Boyd (1980).

bestowed by the physical base. On NRP-2, that is, the powers of the mental are "included" in those of the physical in the sense of being *redundant*. *M*'s non-physical powers are matched by, even if not identical to, some subset of the physical powers bestowed by *P*.[7]

Return now to the problem before us: the first scenario seems not to distinguish reductionism from NRP. In reply, I grant that NRP-1 is empirically indistinguishable from reductionism, even given the assumption of causal intelligibility. But this is not surprising, for this version NRP is, I claim, merely reductionism under another name. Here any causal contribution of the mental property – and indeed the mental property itself – has been entirely absorbed into its physical base. What NRP-1 here calls a "mental property" is just a way of conceiving of a physical base property. Selectively attending to some of *P*'s powers, we single out a referent for our psychological term or concept. But this leaves undisturbed the causal powers of *P* and indeed of any of the powers involved in the production of behavior. Ontologically, NRP-1 is fully reductive: it "adds" a mental property in name only. There is to be sure a "non-reductive" component to NRP-1, but it is merely conceptual or semantic.[8]

However, the alternative version of non-reductive physicalism, NRP-2, does in fact introduce irreducible mental properties, for here *M*'s powers are distinct from any of *P*'s, in spite of the fact that each of *M*'s powers is matched by some physical power in the sense explained earlier. The present question is whether our future investigators, having discovered that mental causation is physically intelligible, will have reason to favor reductionism over this non-reductive rival. When behavior is seen to fit intelligibly with physical causes, is this because, as reductionism would have it, only physical powers are manifested? Or is it because, as NRP-2 would have it, there are two matching sets of powers manifested, one physical and one not?

Distinguishing reductionism and NRP-2 here requires a supplementary assumption, though I believe it is a plausible one. I will call it:

> *Causal Holism*: An effect is unintelligible without all of the powers of which it is a manifestation.[9]

[7] For a defense of NRP-1, see Wilson (2011). For a defense of (what I take to be) a version of NRP-2, see Pereboom (2002).

[8] See Heil (1999, p. 194), as well as Kim's frequent claim (e.g., in Kim, 1992) that the "causal inheritance principle" leads to the fragmentation, and ultimately the reduction, of mental properties.

[9] Compare Williams (2010) on "power holism", a closely related, though not equivalent, thesis.

Given this principle, our empirical investigators, having discovered that mental causation is physically intelligible, will be able to rule out NRP-2. For they would know that if some non-physical power of a mental property were implicated in the production of observed behavior, that behavior would be physically unintelligible. They would know, that is, that on NRP-2, the physiology of intentional behavior should appear as incomplete as a jigsaw puzzle with missing pieces.

Before looking briefly at whether Causal Holism is, or might be, true, I'd like to note as an aside that approaching NRP from this angle allows us to sidestep at least one thorny issue that has occupied much of the mental causation literature. In particular, we need not here decide whether mental properties on NRP-2 overdetermine their effects, and if so, whether such overdetermination is of a problematic sort.[10] If behavior is the manifestation of distinct "matching" powers of mental and physical properties, is this an intolerable coincidence? Is it an unacceptable redundancy in nature? Such questions are important, but I don't think they need to concern us here, for whether NRP-2 should be rejected on empirical grounds in the first scenario is independent of whether it involves a problematic sort of overdetermination. Even if the matching causal powers of *M* and *P* cooperate so as to produce intentional behavior without problematic overdetermination, NRP-2 would still be objectionable after the discovery of physical intelligibility. Our scientists can rule out this non-reductive rival without becoming entangled in metaphysical questions about overdetermination.

That said, ruling out NRP-2 in the first scenario does require Causal Holism, which I haven't here defended – is this just more metaphysical baggage for future scientists to take on in addition to the working hypothesis of causal intelligibility? Perhaps, but it's worth noting that Causal Holism could itself be an empirical hypothesis. Our scientists, investigating causal production in more ordinary (non-mental) cases, could discover that the intelligible production of an effect requires all of the powers manifested, even when those powers include two (or more) with matching causal profiles. This would provide strong, albeit fallible, inductive support for the more general thesis of Causal Holism. Alternatively, a more ambitious strategy would try to establish the principle on broadly philosophical grounds. We might start, for example, by appealing to an analogous holism found in non-causal intelligible structures, such as the jigsaw puzzle or color sphere. A puzzle piece, or a particular shade of red, would not intelligibly fit into a particular location were it not for the natures of the surround-

[10] For helpful discussions, see, e.g., Funkhouser (2002) and Walter (2008).

ing pieces or shades. Such examples, if appropriately analogous to manifestation of powers, could be leveraged into an argument for holism in causal structures. In any case, both lines of argument presented here – empirical and philosophical – are merely suggestive, and a fuller case for Causal Holism would require deeper development of the powers conception of causality, something I will not attempt here.

If these remarks are along the right lines, an empirical discovery – namely, that mental causation of intentional behavior is physically intelligible – would favor reductionism over nonreductive physicalism. I now turn to the other main rival, emergentism. Here the empirical case for reductionism looks more straightforward. After all, the hallmark of emergentism is top-down causation, on which non-physical mental properties exert a novel influence on physical effects, an influence that cannot be accounted for by physical properties alone. Our future scientists, discovering that mental causation is physically intelligible, would be able to rule out emergentism on the spot, and moreover, without needing any principle as controversial as Causal Holism.

But as I noted earlier, there is a version of emergentism which may be compatible with the envisioned empirical discovery. This is the model of mental causation recently defended by E.J. Lowe.[11] Lowe's model has many of the standard emergentist features: mental properties are non-physical, though they depend (causally) on physical properties. Moreover, Lowe's mental properties exert a novel, non-redundant, downward influence on physical (here behavioral) effects. But Lowe's brand of emergentism is unique in that it permits behavior to manifest these top-down powers without violating the closed character of the physical world. Closure has a number of alternative formulations, but here a simple version will do:

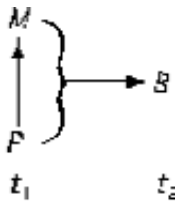
> *Closure*: Everything physical that has a cause at time t has a sufficient physical cause at t.

It's surprising that this principle is compatible with any version of emergentism, for Closure is standardly taken to preclude emergent mental causation. Suppose a behavioral effect B occurring at time t_2 is the manifestation of powers that include those bestowed by mental cause M at previous time t_1. Given Closure, B has a sufficient physical cause at t_1, presumably powers of M's base

[11] For details, see Lowe (2003; 2008). I discuss Lowe's model at greater length in Robb (unpublished paper).

property, *P*. It's difficult to see, then, how *M* could be an emergent property in the required sense. After all, if *M* is non-physical, and its powers are needed to produce *B*, then it looks as if any physical powers operative at t_1 are not sufficient for the behavioral effect after all, so that Closure is violated.

Lowe argues, however, that *M*'s efficacy is in fact compatible with the causal sufficiency of its physical base, and thus with Closure. While the powers bestowed by *M* are necessary partners in the production of *B*, on Lowe's proposed model, *P* is sufficient for *M*. That is, *P* itself produces the needed non-physical property that, together with *P*, produces the behavior. The resulting picture looks like this, where each arrow stands for causal sufficiency:



Here *P* and *M* jointly produce *B*: the mental property is a necessary participant in this production. But Closure is not violated, for *P* is in fact sufficient for *B* due to *P*'s being sufficient for its needed helper, *M*. Lowe remarks that while there is a non-physical property exerting top-down efficacy, it would likely be invisible to empirical investigators, who will find, in accordance with Closure, a sufficient physical cause of the behavior at t_1, and thus will feel no need to postulate non-physical powers.

Lowe's model is ingenious, and deserves more discussion than I can give it here. But for present purposes, the following limited point will suffice: the model is compatible with Closure, but it does not look compatible with physical intelligibility. From the point of view of our empirical investigators, who are operating under the working assumption of causal intelligibility, *P*'s causation of *B* will look extraordinarily odd. We can grant *P*'s causal sufficiency, but because it requires the emergent *M* to produce *B*, *P*'s production of the behavior will look gappy to our scientists, akin to one gear's turning other smoothly in spite of having missing teeth. If this is right, then this first scenario, in which mental causation is discovered to be physically intelligible, will in fact rule out even Lowe's version of emergentism (*a fortiori*, it will rule out more standard versions of emergentism that do not respect Closure). Lowe is right that our future investigators, armed only with Closure, could not empirically decide between reductionism and emer-

gentism. But if they are permitted the hypothesis of causal intelligibility, the case for reductionism under this first scenario looks much stronger.

4. Second Scenario: The Causation of Behavior is not Physically Intelligible

I turn now to a second scenario, one much less friendly to reductionism. Here our advanced physiologists discover that mental causation is *not* physically intelligible. Even after detailed investigation, the causation of intentional behavior appears to them oddly gappy, a puzzle with missing pieces. This would be unacceptable given the working assumption of causal intelligibility. In this case we have what appears to be empirical evidence, not for reductionism, but for one of its chief rivals, emergentism. Mental causation on this second scenario has been found to include non-physical properties bestowing powers not found or matched in the physical world. These non-physical properties are needed to fill in the gaps to make mental causation intelligible.

An immediate worry is that scientists in this second scenario, especially if they already lean toward reductionism, would not have sufficient reason to postulate the powers of emergent mental properties rather than powers of unknown physical properties, or, alternatively, unknown powers of familiar physical properties, powers manifested only in psychological contexts.[12] Reductionism remains an option for them, and the second scenario doesn't vindicate emergentism after all. This particular worry, however, misrepresents the current dialectic. I am here concerned only to show, under the assumption of causal intelligibility, the in-principle empirical resolution of the mental causation debate. To secure this, we can suppose in these hypothetical scenarios that the relevant empirical information about the contribution of physical powers to behavior is available to our future physiologists. Skeptical worries about physical properties and their powers are set to one side, at least for present purposes.

A related and more troublesome worry about the second scenario asks why our scientists should think the non-physical properties in question are *mental*. Perhaps mental properties are physical, as the reductionist would have it, and the top-down powers filling in the gaps of mental causation are, while not physical, not mental either. This is not as *ad hoc* a hypothesis as one might think. If there is reason to countenance emergence in non-psychological contexts – perhaps chemistry could provide examples[13] – these non-mental emergent powers could

[12] On this latter possibility, compare Shoemaker (2002).
[13] See Hendry (2006).

be part of the necessary causal background in which mental causation operates. Our scientists in the second scenario might have discovered emergent causation, but perhaps not emergent *mental* causation. To respond to this worry, we need, I think, to supplement the second scenario so that the efficacy of non-mental emergents, if there are such, is already taken into account when gaps are perceived in the causation of behavior. The key here is for our future scientists to study these non-mental emergents outside of psychological contexts. Constructing causal profiles for these properties, scientists can take their efficacy into account when investigating the causal history of behavior. Only then could failures of intelligibility plausibly be attributed to emergent *mental* properties.

So far the second scenario appears to be one in which there is evidence for emergentism over reductionism. But as in the first scenario, the difficult rival to rule out is NRP-2. Our scientists see that behavior is not physically intelligible, and so under the assumption of intelligibility, they can infer that this is due to the operation of non-physical powers. But are these postulated powers novel, as emergentism has it, or are they the redundant powers of NRP-2? Causal Holism will not be of much help in this case. Holism tells us that an effect will appear unintelligible until we grasp all of the powers producing it, but the second scenario already presents the causation of behavior as physically unintelligible, and thus as involving hidden (non-physical) powers. The question is not whether there are hidden powers, but what their natures are. In particular, are they novel or redundant with respect to the physical? If there is empirical headway to make on this question, it will not come from constructing casual profiles for the postulated non-physical powers in non-mental contexts, for they do not occur in such contexts. If there's to be empirical evidence deciding between NRP-2 and emergentism, it will rather have to come from studying the intelligible manifestation of ordinary (non-mental) powers under conditions of *known redundancy*. If a telltale sign of redundancy could be gleaned from such cases, it could shed light on the nature of the hidden powers in the second scenario. Knowing how redundancy reveals itself, our scientists could leverage this information to decide between emergentism and NRP-2.

5. Conclusion

The mental causation debate is as much about how mental causation occurs as it is about whether it does.[14] The "whether" question is important, but I've set it

[14] See Kim (1998, pp. 61–26, pp. 78–79).

aside to focus on the "how". In particular, how are the mental powers producing intentional behavior related to the physical, biochemical powers observed to work in the human body? *Prima facie*, these are empirical questions, and accordingly there have been attempts to marshal empirically confirmed – or at least empirically plausible – principles to bring the mental causation debate to some resolution.[15] For example, perhaps we have good empirical reasons to think that Closure is true, or that energy is conserved, or that causation is the transference of some measurable quantity, or that there must be an underlying mechanism for any macro-level causal relation. Armed with such principles, we can look to a sufficiently advanced physiology to tell us whether the physical world has the resources to explain the causation of behavior. But the availability of NRP-2, as well as Lowe's "invisible" model of emergent causation, suggests that even advanced empirical evidence will be of limited use in resolving the mental causation debate, at least when only Closure and kindred principles are in play. My aim in this paper has been to argue that the working hypothesis of causal intelligibility would make empirical progress on these questions a more reasonable hope.

REFERENCES

Boyd, R. (1980). Materialism without Reductionism: What Physicalism Does Not Entail. In N. Block (Ed.), *Readings in Philosophy of Psychology*, Vol. 1. Cambridge, MA: Harvard University Press.

Chalmers, D. (1996). *The Conscious Mind*. Oxford: Oxford University Press.

Funkhouser, E. (2002). Three Varieties of Causal Overdetermination. *Pacific Philosophical Quarterly*, 83, 335–351.

Heil, J. (1999). Multiple Realizability. *American Philosophical Quarterly*, 36, 189–208.

Heil, J. (2003). *From an Ontological Point of View*. Oxford: Oxford University Press.

Heil, J. (2012). *The Universe as We Find It*. Oxford: Oxford University Press.

[15] For selective surveys of such attempts, and from quite different perspectives, see Lowe (1992) and Papineau (2001).

Hendry, R.F. (2006). Is there Downward Causation in Chemistry? In D. Baird, E. Scerri, & L. McIntyre (Eds.), *Philosophy of Chemistry: Synthesis of a New Discipline; Boston Studies in the Philosophy and History of Science*, 242, 173–189.

Kim, J. (1992). Multiple Realization and the Metaphysics of Reduction. *Philosophy and Phenomenological Research,* 52, 1–26.

Kim, J. (1998). *Mind in a Physical World*. Cambridge, MA: MIT Press.

Lowe, E.J. (1992). The Problem of Psychophysical Causation. *Australasian Journal of Philosophy,* 70, 263–276.

Lowe, E.J. (2003). Physical Causal Closure and the Invisibility of Mental Causation. In S. Walter, & H. Heckmann (Eds.), *Physicalism and Mental Causation*. Imprint Academic.

Lowe, E.J. (2008). *Personal Agency*. Oxford: Oxford University Press.

Mumford, S., & Anjum, R.L. (2011). *Getting Causes from Powers*. Oxford: Oxford University Press.

Papineau, D. (2001). The Rise of Physicalism. In C. Gillett, & B. Loewer (Eds.), *Physicalism and Its Discontents*. Cambridge: Cambridge University Press.

Pereboom, D. (2002). Robust Nonreductive Materialism. *Journal of Philosophy*, 99, 499–531.

Putnam, H. (1980). The Nature of Mental States. In N. Block (Ed.), *Readings in Philosophy of Psychology*, Vol. 1. Cambridge, MA: Harvard University Press.

Robb, D. (2007). Power Essentialism. *Philosophical Topics,* 35, 343–358.

Robb, D. (forthcoming). Power for the Mental as Such. In J.D. Jacobs (Ed.), *Putting Powers to Work*. Oxford: Oxford University Press.

Robb, D. (unpublished paper). Could Mental Causation Be Invisible?

Sanford, D.H. (1975). Causal Necessity and Logical Necessity. *Philosophical Studies*, 28, 103–112.

Sanford, D.H. (1994). Causation and Intelligibility. *Philosophy,* 69, 55–67.

Shoemaker, S. (2002). Kim on Emergence. *Philosophical Studies*, 108, 53–63.

Walter, S. (2008). The Supervenience Argument, Overdetermination, and Causal Drainage: Assessing Kim's Master Argument. *Philosophical Psychology*, 21, 673–696.

Williams, N. (2010). Puzzling Powers: The Problem of Fit. In A. Marmodoro (Ed.), *The Metaphysics of Powers*. New York: Routledge.

Wilson, J. (2011). Non-reductive Realization and the Powers-based Subset Strategy. *The Monist*, 94, 121–154.