13

Nonreductive Physicalism or Emergent Dualism? The Argument from Mental Causation

Timothy O'Connor and John Ross Churchill

Throughout the 1990s, Jaegwon Kim developed a line of argument that what purport to be *nonreductive* forms of physicalism are ultimately untenable, since they cannot accommodate the causal efficacy of mental states. His argument has received a great deal of discussion, much of it critical. We believe that, while the argument needs some tweaking, its basic thrust is sound. In what follows, we will lay out our preferred version of the argument and highlight its essential dependence on a causal-powers metaphysics, a dependence that Kim does not acknowledge in his official presentations of the argument.[1] We then discuss two recent physicalist strategies for preserving the causal efficacy of the mental in the face of this sort of challenge, strategies that (ostensibly) endorse a causal powers metaphysics of properties while offering distinctive accounts of the physical realization of mental properties. We argue that neither picture can be satisfactorily worked out, and that seeing why they fail strongly suggests that nonreductive physicalism and a causal powers metaphysics are not compatible, as our original argument contends. Since we also believe that robust realism concerning mental causation should not be abandoned, we take the argument of this paper to strongly motivate an account on which the mental is unrealized by and ontologically emergent from the physical. In a final section, we sketch what an ontologically emergentist account of the mental might look like.

[1] While Kim does not officially endorse a causal powers metaphysic, he has noted his reliance upon a certain view of causation in making his case, a view that, on the surface, bears some similarity to the view we endorse. See for example Kim (1998: 45–56); Kim (2002: 674–5); and Kim (2005: 17–18, 30, 45, and 47 note 12). However, there is some ambiguity in the way Kim uses the term 'cause.' See Kim (1998: 43) and Kim (2005: 20 and 41). And see especially Kim (2002: 677), where he speculates that causality may supervene on fundamental laws (and, perhaps, initial conditions), or that it could 'emerge' at macro-levels ('Could Hume be right about fundamental physics but wrong about macro-objects and events?'), or that it might be 'implemented' or 'realized' by something more basic, like energy flow or momentum transfer.

1. A CAUSAL POWERS ONTOLOGY

Let us first explain what we mean by the term 'causal powers.' One way of using this term is merely a loose manner of describing the causal features of a property, entity, or kind independently of any definite commitments on the metaphysics of causation. A person using the term this way might say, for example, that a defoliant has the causal power to kill plants, where this claim is neutral as to whether (a) the presence of the defoliant can produce or generate the death of a plant; (b) there is a law of nature that relates properties of the defoliant and plant death; (c) plants regularly die after being sprayed with the defoliant; (d) there are subjunctive conditionals relating the properties of the defoliant and the death of plants in a certain way; (e) citing the presence of the defoliant satisfactorily explains the occurrence of plant deaths in certain contexts; or some (f) fitting a still different analysis of causation.

We do not use the term in this neutral manner. Our usage corresponds to the first of these: a power to produce or to generate, where this is assumed to be a real relation irreducible to more basic features of the world. Our favored technical term for this is 'causal oomph.' So understood, causation is not amenable to analysis in non-causal terms, but instead involves the exercise of ontologically primitive causal *powers* or *capacities* of particulars. Powers are either identical to, or figure into the identity conditions of, certain of the object's properties, which are immanent to those things as non-mereological parts. (Whether one thinks of these as immanent universals or tropes is not crucial in this context.)

It bears emphasis that this view is *not* committed to assuming that all causation must amount to something like 'pushing,' or 'pulling,' or 'knocking,' or 'the exertion of a force.' What is assumed, rather, is solely this: when an instance of a property—the event of the particular's having the property—is a cause, the world unfolds in a certain way after the instance of that property, and that property instance is one of the factors that jointly *make* the world unfold this way. This is just another way of saying what's come before, that the property instance, and others besides, jointly *produce* or *generate* certain effects; they jointly oomph the world into going on in *this* way rather than *that*. Because of this, there are certain counterfactuals true of the world ('were the property not to have been instanced, such-and-such effects would not have occurred'). But these counterfactuals are derivative from, and not to be equated with, or seen as the basis of, the causal facts themselves: it's *because* the property instance was among the factors that jointly produced the relevant happenings that certain corresponding counterfactuals are true. Causally efficacious properties have the power to make the world unfold in ways that it otherwise would not, and this is a fundamental feature about these properties upon which all else (counterfactuals true of them, regularities and patterns that encompass them, explanations that cite them) is derivative.

There is much debate, and not a little confusion, over how to delineate the finer points of this general picture. While we cannot delve deeply into these matters, we make the following two remarks to forestall confusion that might infect understanding of our subsequent argument. First, there is a pervasive manner of speaking that appears on the surface to say that *objects* have and exercise causal powers. (Witness our example above with respect to defoliants.) In our view, such talk should be construed by the causal powers metaphysician as a shorthand way of expressing the claims that:

i) the having of the property is the having of the causal power;
ii) the event of the property's being had by the object in appropriate circumstances causally contributes to the effect; and
iii) the exercise of the causal power just is this causal contribution.[2]

Second, a single property may contribute to a very wide array of effects, depending on the context in which it is instanced. A particle's being negatively charged may contribute to its accelerating at varying rates away from a similarly charged nearby particle, accelerating toward an oppositely charged nearby particle, even accelerating towards a similarly charged particle (though at a slower rate than would occur were the particle not to have been so charged), and countless other manifestations, all depending on the context of its occurrence. But in ordinary speech, again, there is a tendency to talk of a corresponding array of causal powers being exercised, 'each' of which is identified through the effect actually manifested. This sort of speech has encouraged some metaphysicians to posit a multiplicity of properties, or worse, to posit a distinct type of entity (a causal power), any number of which are 'conferred by' a single property. We should resist such moves on grounds of parsimony, and here science is a much better guide to property/power identifications.[3] The key is to understand a basic power or disposition not in terms of this or that salient manifestation, but rather in terms of a unitary causal influence, something that is constant across circumstances while its manifestations will vary.

While we cannot undertake here a defense of this understanding of causation, we will summarize what we take to be some key advantages over two very general rival approaches. The first is the class of broadly Humean reductionist accounts. While details differ considerably, on all Humean accounts, whether one event

[2] One *might* hold to a philosophical view leading one to insist that in certain cases, it is indeed the object that exercises the power, and not the event of the object's having the property/causal power. Such is the claim of the agent causationist, e.g., with respect to the forming of a free decision. But this is a substantive and controversial thesis, not a spelling out for one sort of case what is common to every case of causation. (For a discussion of the relationship of agent causation to the more usual 'event causation' within a causal powers metaphysics, see Timothy O'Connor, 'Agent-Causal Power,' forthcoming in Toby Handfield (ed.), *Dispositions and Causes* (Oxford: Oxford University Press).

[3] We are influenced here by Richard Corry. See 'A Causal-Structural Theory of Empirical Knowledge' (PhD thesis, Indiana University, 2002) and 'Scientific Analysis and Causal Influence,' in Handfield (2008).

causes another is a massively nonlocal matter, insofar as causal relations supervene on the global pattern of events across space and time. Should the pattern of future events turn out to be very different from what our best theories now predict, it might 'turn out' that what we thought to be an obvious causal interaction—never mind the details of its nature, the very existence of any causal relationship at all—was no such thing. But this is implausible. What happens in the distant reaches of space and time cannot nullify whether a causal transaction occurs here and now. (To focus, one's intuitions, consider rational agency as a special case. Actions are a special kind of causal process. If a Humean picture of causation more generally is correct, then should the future pattern of events unfold in certain ways, and nothing that has happened thus far will keep it from doing so, we should have to say none of us ever undertook an action at all. Mind, we are not speaking here of 'free' actions in the sense of free will, just plain-old actions.) The causal powers metaphysician has no such implausible nonlocality commitment. Furthermore, by grounding causal relationships, ultimately, in the contingent global distribution of noncausal facts, certain kinds of explanations become unavailable in principle. For example, if, in seeking an explanation for the occurrence of token event y, we're seeking knowledge of *what made y occur*, then the Humean must deny that there is any such item of knowledge. And while, e.g., X- and Y-type events may conform to a pervasive pattern of a specified sort (whether actual or counterfactual), for Humeans there will either be no explanation as to why that pattern holds, or else the explanation will itself bottom out in unexplained pattern facts. For anyone who shares our temperament vis-à-vis explanation, these consequences are bound to disappoint.[4] The account of causation we favor, on the other hand, invites neither of these two disappointments. This should be obvious in the first case, given what we have proposed concerning the causal relation. As for the second case, noting that X-type events all manifest a common property that is disposed to bring about Y-type events in specified sorts of circumstances provides us with an explanation as to why X- and Y-type events conform to the pattern they do, an explanation that does not bottom out in unexplained pattern facts. (And whatever else we might think of them, Humean complaints that dispositions introduce a 'mysterious modality' into the world are hardly founded in scientific practice, where functional methods of specifying theoretical properties is standard. On the causal powers view, where such identifications are accurate, they capture something about the nature of the property itself. Whereas for the Humean, they merely describe contingent patterns of instantiation of the property, while being forever silent concerning its intrinsic character.)

The second rival approach, developed in somewhat different ways by Michael Tooley and David Armstrong, involves associating particular causal relations in

[4] We are aware that not everyone shares our temperament with respect to explanation, so not everyone will share our disappointment. For more on causal explanation, to include thorough discussion of extant theories of causal explanation, see Woodward (2003).

the world with a contingent, higher-order, nomic relation among universals, a relation that Armstrong dubs 'necessitation.' The central idea is that the obtaining of a nomic fact of the form **N** (**F**, **G**) grounds and explains the fact that a particular instance of **F** is followed by a particular instance of **G**. But, as many have pointed out, it's not at all transparent how a second-order relation among universals can constrain particular first-order **F**-**G** sequences.[5] Both Tooley and Armstrong have tried to complicate the account in order to overcome the difficulty. Here we restrict our attention to Armstrong's (1997) strategy. It is to conjecture that the second-order relation among universals is identical to the causal relation among its instances—causation is a relation not among particular, first-order states of affairs, but of types of states of affairs. But now the appearance of an advantage over the brute conjecture of the Humean theorist has vanished. For each occurrence of **G** is ontologically and so explanatorily prior to the immanent, co-occurring **N** (**F**, **G**) fact. So the posit of the **N** relation is gratuitous, as it can only be put into the world consequent upon the regularity. It is merely a baroque adornment to Humeanism—enough so that we might with justice call it 'second-order Humeanism.'[6] This is so, that is, unless we make the stronger claim that **F** by its very nature is disposed to bring about **G**, in which case we are back to the primitive dispositionality of the causal powers metaphysics.

Such, in outline form, are a few central reasons we have for thinking a causal powers metaphysics to be preferable to its main rivals. In considering the prospects for a nonreductive physicalist view of the mental, we are assuming, rather than arguing for, this causal powers metaphysic. We are investigating its implications for the question at hand. Can the (by our lights) right-thinking metaphysician who has seen his way clear to this view of causation make out a nonreductive physicalist view on which mental states are causally efficacious in this sense? We will try to persuade you that the prospects are bleak.

2. CAUSAL POWERS AND THE DILEMMA OF REDUCTION OR CAUSAL EXCLUSION OF THE MENTAL

We will now present our preferred version of a Kim-style argument for the reducibility of mental properties to physical properties.[7] We begin with three related premises concerning causation and properties:

1) Causation is a real relation irreducible to more basic features of the world. *(causal nonreductionism)*

[5] See van Fraassen (1988: chapter 5), and Lewis (1983a: 40).

[6] Mumford (2004) includes an extended discussion of this problem for the Armstrongian approach. See pp. 99–103 and 148–49.

[7] See Kim (1993b, 1997, 1998, 1999, 2003, 2005).

2) Causation involves the exercise of causal powers or capacities of particulars. *(production account of causation)*

3) Properties are individuated in terms of causal powers, such that there are no distinct properties that confer exactly the same causal powers. *(causal theory of properties)*

The next three premises flow from the distinctive commitments of non-reductive physicalists:

4) Mental properties supervene on physical properties. *(supervenience thesis)*

The hoary slogan, of course, is 'no mental difference without a physical difference,' intended to capture an appropriate dependence relation. What exact form the supervenience relation should take in this context, however, is a difficult and controverted issue. We will follow Kim in supposing that complication arising, e.g., from mental content externalism can be safely ignored. If this is correct, we may assume for the sake of argument that mental properties 'strongly supervene' on the physical properties of the individual (or on the physical properties and relations of the individual's parts). Next we have:

5) Mental properties are realized by physical properties: a particular event *M* of a person *S*'s having mental property **M** is either 'constituted by' (a kind of ontological posteriority) or is identical to various physical particulars—possibly including portions of the person's environment—having certain physical properties and standing in certain physical relations. *(realization thesis)*

We will be noncommittal on whether the realization of mental properties by physical properties involves constitution or identity of the corresponding events, since non-reductive physicalists' pronouncements on this matter are varied and often obscure.[8] Finally, physicalists typically wish to assert:

6) Every *physical* event that has a cause has a complete physical cause.[9] *(causal completeness of physics)*

According to (6), nothing non-physical is *required* in order to causally account for the occurrence of any physical event, where the latter consists in the instantiation of fundamental physical properties and relations by fundamental physical entities. Whatever their particular views concerning the status of special science laws and causes, including those pertaining to psychology, the typical physicalist maintains that any fundamental physical event (including large-scale distributions of fundamental properties and relations) that has a cause has a cause that is equally fundamental and physical in character. The true physics is causally complete.

[8] See, for example, the variation among Fodor (1974); Pereboom and Kornblith (1991) and Pereboom (2002); Shoemaker (2001 and 2007); and Gillett (2002).

[9] We will ignore the complication of indeterministic causation, which would require us to formulate the completeness thesis in terms of fixing the chances of the effect.

We now contend that (1)–(6) are inconsistent with supposing

7) A mental property, **M**, is distinct from its physical realizer property (or properties), **P**, and each event that consists of **M**'s being instanced exercises a distinctive form of causality that one way or another impinges the realm of physical events.[10] *(assumption for reductio)*

Premise (7) is the supposition that there are mental properties that do not reduce to physical properties and whose causal efficacy does not reduce to the causal efficacy of some physical properties. This means, in the schema used in (7), that the singular causal action of the mental event of **M**'s being instanced does not reduce to the singular causal action of some physical event or events (say, the instancing of the physical property **P** that realizes **M** in the circumstances). In short, the commitment expressed by (7) is what puts the 'non' in 'non-reductive' physicalism.

The argument that (7) is inconsistent with (1)–(6) proceeds as follows.

8) The instance of **M** either
   (a) directly produces a subsequent mental event, M^*, or
   (b) it directly produces a wholly physical event, P^*.

The realization thesis (5) and production account of causation (2) together strongly suggest that option (a) is a nonstarter. On this view, mental events are ontologically dependent on their subvening realizers, wholly constituted by (if not identical to) them, and this is no less true of mental *effects* as of mental *causes*. Bringing about such a mental event *eo ipso* involves causally affecting the physical event which realizes it. So

9) Not (8a).

But the thesis of causal completeness (6) implies that

10) If 8b, then the physical event P^* is overdetermined by M and some other physical event.

Now, if we accept the production account of causation, it will seem passing strange to suppose that, in regular fashion, there are physical events that are systematically 'overoomphed' by distinct events, even if—indeed, *especially* if—these causes might stand in a supervenience relation. If, say, a physical event P, the realizer of the mental event M, produces or oomphs P^*, what causal work is left over for M? It would be at best a gross violation of parsimony to posit two distinct productive relations for a single event every time mental events supervene

[10] We will, for the sake of convenience, continue to refer only to **P**, the single realizer of **M**, though it should be understood that on some accounts of realization **M** may be realized by multiple properties ('the **P**s,' say) each time it is instanced. As we'll see below, Gillett (2002) is one such account.

on the fundamental physical cause. Note that on reductive accounts of causation, on which causal facts are not something additional to the totality of noncausal facts, the situation looks very different. Suppose, for example, that our effect P^* is counterfactually dependent on both P and M. If we accept something like the counterfactual analysis of causation, there is nothing strange or objectionable about deeming M, as well as P, to be a cause of P^*. For in doing so we are not making a commitment to anything additional—M's status as a cause of P^* falls out of the facts that we already accept along with our analysis. It comes for free. By contrast, on the nonreductive productive account, we would be positing an additional fundamental relation between M and P^*, when doing so is entirely unnecessary for accounting causally for P^*. Thus, we should conclude that:

11) There is not systematic mental-physical overdetermination, as the consequent of 10 implies.

But this is the end of the road. We are forced to conclude, therefore, that:

12) **M** does not make a distinctive contribution to occurrences in the physical world, whether wholly physical or supervening mental occurrences. *(completing Reductio of 7)*

Finally, the causal theory of properties (premise 3) both rules out an epiphenomenalist retreat and suggests the proper ultimate conclusion: we ought either to reductively *identify* **M** with **P** or *deny* that **M** is a bona fide property—one that earns its causal keep—in the first place.

Nonreductive physicalists see an obstacle to the first option, reductionism, in the fact that, as functional properties, intentional properties are multiply realized. What counts as a belief that Q in humans may be quite distinct, at any physical level of description, from what counts as that same belief in, say, an intelligent extraterrestrial or a sophisticated artificial machine built out of steel and silicon. In reply, Kim recommends that we seek local, species-specific reductive identities for intentional properties—*human* belief that such-and-such as identical with physical property so-and-so—and so preserve the status of these intentional properties as causal powers. That is, we characterize both **M** and **P** in terms of *highly* specific mental and physical types, respectively, and move to a type-type identity theory.

The second, eliminativist option is to interpret apparent reference to mental properties as properly denoting mental *concepts* only. There are far fewer properties had by an object than the vast number of concepts it falls under. As indicated in premise (3), *properties* are immanent to their instances and make a nonredundant difference to how the objects act in at least some circumstances. (They answer to what Kim calls 'Alexander's dictum.')

The argument just presented, like earlier relatives, seeks a reductionist or eliminativist conclusion by way of arguing for the *exclusion* of irreducibly mental causation. Yet it does this by explicitly invoking the thesis of causal powers

realistically construed. So let us refer to it hereafter as the *power exclusion argument*.

As critics of Kim have observed, this argument appears to generalize beyond mental properties to all properties posited in the special sciences (sciences other than basic physics).[11] And since, contra Kim, it is highly plausible that special science categories are not ontologically reducible (owing in part to their own multiple realizability),[12] the eliminativist conclusion the argument ultimately invites here is often taken as a *reductio ad absurdum*: surely the terms of well-established biological and chemical theory pick out genuinely efficacious properties!

Owing to length constraints, we shall not be able to treat this sort of indirect criticism of the argument in detail. We will rest content with the following two-fold response.

First, notice that a rejection of premises (5) and (6), the realization and causal completeness theses, suffices to block the final conclusion of the power exclusion argument. As we discuss later on, we believe the best way to maintain a robust, nonreductive view of the mental is to reject these two premises. Similarly, one might reject the corresponding premises in an exclusion argument directed at special science properties that one takes to be irreducible and efficacious. Recent philosophy of science has seen significant challenge to the completeness thesis in particular.[13]

But second, for one who takes the case for the completeness of physics with respect to some or all special sciences to be convincing, it would not be absurd to accept a causal exclusion conclusion from a corresponding form of argument. For so-called 'high level' theories can be enormously useful and illuminating, and even necessary to the progress of human knowledge of how the world works, without answering to ontological 'levels' or layers populated by distinctive properties and their objects.[14] And the further fact that such theories are not generally reducible to more fundamental theories is a highly interesting fact about our world (and one necessary for science to get off the ground, as in practice we inevitably work our way in, not out), but it cuts no ontological ice. An alternative to the levels picture of physical reality has already been hinted at above: there is a vast array of microphysical entities (for simplicity, 'the particles') bearing primitive, dynamical features and standing in primitive relations. Talk of composite objects and their properties, at least in the general case, is the imposition of a conceptual scheme that selectively picks out coarse-grained patterns running through the vast storm

[11] For discussion, see Baker (1993); Burge (1993); van Gulick (1993); Kim (1996, 1997, 1999, 2003, 2005); Block (2003); Ross and Spurrett (2004).

[12] See Fodor (1974); Dupré (1993); and Rosenberg (1994).

[13] See Cartwright (1999) and Dupré (1993 and 2001). And for a powerful challenge to the case for completeness in the special scientific domain of chemistry in which it is widely thought to be most secure, see Hendry (2006).

[14] On this point, see Heil (2000: chapters 2–7).

of particles. These concepts really are (*objectively*) satisfied by the world, but not in virtue of a one–one relation between general concepts and properties, or individual concepts and particulars.

This second response might be thought to entail the devaluation of the special sciences. Such a conclusion would be too hasty, however. For it is simply false that science is of value only as a source of representing the world in more and more accurate ways. It is, in addition, a source of means for intervening and manipulating the world so as to change it for the better, and much of its value is due to this rather than its representational fruits. We value science—we fund it, prioritize it, give special social status to many of its practitioners, etc.—because of its role in improving the world, and not just because of its role in representing the world. (The development of methods for effectively preventing and treating myriad diseases serves as just one example of such improvement.) But *qua* sources of improvement, some of the special sciences are at least as valuable, and perhaps more so, than fundamental physics. For we are very often *better* able to intervene and manipulate in ways that improve the world by using the resources of the non-fundamental special sciences.

Returning to the status of mental properties, here, then, is where we are left. The commitments that drive the power exclusion argument—the causal powers metaphysics along with the supervenience, realization, and causal completeness theses of nonreductive physicalism—appear to generate the conclusion that mental properties are either reducible or eliminable. This is serious trouble for philosophers who are neither reductionists nor eliminativists with respect to the mental, and they are legion. Thus, if we wish to preserve the mind as irreducibly efficacious, we must reject one or another of the commitments driving the argument above.

Or so we believe. Sydney Shoemaker, however, disagrees, and has recently attempted to provide a way out for the nonreductive physicalist who is a realist with respect to causal powers. Since Shoemaker has *bona fides* as both a causal powers metaphysician and as a physicalist, it is fitting that we investigate his approach in detail.

3. SHOEMAKER ON NON-REDUCTIVE MENTAL CAUSATION

Shoemaker thinks that the key to vindicating the causal efficacy of mental properties without reduction lies in a distinctive account of the *realization* of mental properties by physical properties. In broad strokes, his proposal is that mental and other realized properties are *disjunctive* properties, with their disjuncts as their realizers: the relation of realizer to realized is simply the relation

of disjunct to disjunction.[15] This ensures that realized properties have a proper subset of each of their realizers' *forward-looking causal features*—what instances of the properties can causally suffice for—while having a superset of their realizer properties' *backward-looking causal features*—what can causally suffice for instances of the properties. Shoemaker then exploits the conclusion that realized properties have a subset of their realizers' powers to argue that mental causation is not reducible to causation by the physical realizers, owing to a certain proportionality thesis explained below concerning what counts as a cause of what. The following schema captures Shoemaker's picture.

(**C1 v C2**) is a property realized by each of **C1** and **C2** just in case:

$C1 \rightarrow E1$	$C1 \rightarrow (E1 \text{ v } E2)$	$B1 \rightarrow C1$	$B1 \rightarrow (C1 \text{ v } C2)$
$C2 \rightarrow E2$	$C2 \rightarrow (E1 \text{ v } E2)$	$B2 \rightarrow C2$	$B2 \rightarrow (C1 \text{ v } C2)$
	$(C1 \text{ v } C2) \rightarrow (E1 \text{ v } E2)$		$(B1 \text{ v } B2) \rightarrow (C1 \text{ v } C2)$

('→' denotes causal sufficiency)

How is accepting this picture of realization supposed to make things easier for non-reductive physicalism? We begin by observing that if the realized property has a subset of the forward-looking causal features of the realizer, then the realizer property event is causally sufficient for everything the realized property event is causally sufficient for, *plus more*. So, for example, *C1* is causally sufficient for (*E1* v *E2*), just as (*C1* v *C2*) is, but unlike the disjunctive cause it is also sufficient for an instance of **E1**. Now, if *C1* and (*C1* v *C2*) overlap in this way in what they causally suffice for, and if causal considerations ought to drive our conclusions about the identity of properties, a natural conclusion is that (*C1* v *C2*) is a proper *part* of *C1*. More generally: events which instance realized properties are parts of those instancing the corresponding realizers, and so are not identical to them.

From here, Shoemaker invokes a version of Stephen Yablo's 'proportionality' constraint[16] on what we ought to count as 'the cause' in a causal interaction: while it is true that *C1* is causally sufficient for (*E1* v *E2*), (*C1* v *C2*) is, Yablo and Shoemaker say, a *better candidate* for being the cause. For (*C1* v *C2*) is also causally sufficient for the specified effect, but only 'just so'—it causally suffices for the effect *and nothing more besides*. The only features of *C1* that contribute to the 'bringing about' of (*E1* v *E2*) are features had by (*C1* v *C2*), a 'part' of *C1*. And

[15] For key passages in support of this interpretation, see Shoemaker (2007: section II of chapter 2 (especially pp. 17–18), pp. 55–6, and section V of chapter 4 (especially p. 79 and 82)). See also the remark that Lenny Clapp has proposed a view similar to his own in Shoemaker (2001: 93 note 3, and 2007: 11). (Clapp (2001) is explicit in his construal of realized properties as disjunctive properties with their disjuncts as the realizers.)

[16] Shoemaker (2001: 81). See Yablo (1992).

as with a more familiar sort of case, such as Jones's single shot in a firing squad, just ahead of the others, killing the condemned, we are invited to conclude that it is best to say that while the whole (*C1*; the firing squad's firing) was causally sufficient for the effect ((*E1* v *E2*); the death of the condemned), proportionality constraints argue in favor of counting a particular part ((*C1* v *C2*); Jones's firing) rather than the whole as *the cause*.[17] This is how realized events in general—and realized *mental* events in particular—qualify as causes in certain scenarios, such that it is *false* in these scenarios that the (physical) realizer events are likewise causes of the very same effects.

We now have Shoemaker's account of realization laid out before us, as well as the way it is supposed to provide for non-reductive mental causation. But how, we may wonder, does the account underwrite a response to the power exclusion argument? Notice, first, that there is no rejection of the supervenience, realization, or completeness theses (premises 4–6). And, of course, Shoemaker intends (7), the anti-reduction premise, to come out true as well. Thus, one who takes our exclusion argument above to be cogent will naturally suspect that Shoemaker's commitment to the causal powers metaphysics (as expressed in premises 1–3) is less than it appears.

To bring the problem into focus, consider first that, for all his distinctive claims, Shoemaker clearly gives ontological priority to the physical realizer event. He tells us that *P* realizes *M* just in case *P* is metaphysically sufficient for (but not identical to) *M* and 'constitutively makes [it] real' (Shoemaker 2007: 6, 10). He goes so far as to say that realized states are 'nothing over and above' their realizers (Shoemaker 2007: 2). If all this is so, then how is a case of *M*'s causing an effect, *E*, not also a case whereby *P*, *M*'s constituting realizer, is likewise causing *E*? Indeed, how is this not a case where *P* is causally prior to *M*, so that, by the power exclusion argument, we should conclude that *P* is the sole true cause?[18] It seems that only a retreat from a causal powers metaphysics could allow you to say that *P* is somehow 'merely' causally sufficient whereas *M* is the proper cause. If *P* is ontologically prior to *M*, *able* to bring about *E*, and in the circumstances necessary to do so, how can it get out-oomphed by *M*?

The only way for us to make sense of this is by ignoring Shoemaker's talk of *P*'s constitutively making real *M* and focusing instead on his notion that *M* is a part of *P*, owing to the subset-of-powers thesis and the causal theory of properties. (This line of interpretation is encouraged by his invocation of the

[17] See Shoemaker (2001: 81) and (2007: 13–14).

[18] A bolstering consideration comes from certain indeterministic scenarios. We take it to be evident that, *assuming the causal completeness of physics*, the chance of *E* given *M* cannot be greater than the chance of *E* given a total physical cause (here, our *P*). But there seems to be no reason to think that it cannot be less. Now consider a case where Pr (E/M) is significantly less than Pr(*E*/*P*). Surely in such a case, where *E* in fact occurs, it is highly implausible to insist nevertheless that *M*, not *P*, is the cause of *E*. While this is a special case, if our conclusion from it is accepted, it indicates further that there is something wrong about Shoemaker's method for assigning causes.

firing squad analogy (Shoemaker 2007: 53) and also by such statements as 'It is only because the C-fiber stimulation instance realizer contains the pain instance realizer that it has the relevant effects.' (Shoemaker 2007: 48)) We might then suppose that Shoemaker is proposing what amounts to a radical inversion of the reductionist's vision, such that it is the *physical* properties that resolve into an assemblage of mental properties plus some non-mental remainder.[19] Now, even if that could make sense of how mental causation would be genuine, it has the substantial drawback that it is just plain weird. The physical property is thereby conceived as (or as closely bound up with) a cluster of causal features, a subset of which are the features that define an associated mental property. Making this picture out perhaps requires us to analyze the physical property as a structural property, the instancing of which just consists in the instancing of properties of the object's parts and relations between them. The mental property then comes out, on Shoemaker's view, as an overlapping structural property, perhaps somehow abstracted from the full physical structural property. Waiving considerable doubts we should have about the plausibility of this picture of the mental-physical relationship, it still has the result that *both* mental properties and the larger structural physical properties in which they are embedded turn out to be derivative structures, entities that are constructions out of *micro*physical properties and relations. The specter of reductionism again menaces.

Shoemaker attempts to resist the reductive identification of mental (or other macro-level) properties with microphysical states of affairs by giving an analysis of microphysical realization on which there is only constitution, not identity. However, his case for this rests on two claims about property identity that should be unacceptable to a causal powers theorist (Shoemaker 2007: 48–9). First, he lays down that, in general, a property instance has just one constituent object and one constituent property, so a mental property instance can't be identical to a state of affairs involving many distinct properties and objects. He seems to put this forward as a definitional truth or platitude. But a causal powers theorist does not take quasi-grammatical considerations to be final arbiters concerning the structure of reality. One might just as well take Shoemaker's supposed platitude together with facts (assumed for now) about microphysical constitution and draw the conclusion that there are not, strictly speaking, mental properties at all—not in the sense of entities that contribute directly to how the world unfolds. Shoemaker's second claim is that the modal properties of macro-level property instances and their microphysical realizers will generally differ. (Consider familiar claims made in discussions of the statue of Goliath.) This claim rests on intuitive judgments about possible variation in the material constitution of composite objects. But the status of composite objects, no less than that of 'their' properties, is very much in question on the powers metaphysics. One cannot simply assume

[19] We've recently come across a paper where a similar interpretation of Shoemaker is entertained. See Heil (2003: 24).

that there are robustly *objective* modal facts about them and use these to ward off what otherwise appears to be a powerful reductionist challenge.

It is time to take stock. We have defended a power exclusion argument for the untenability of nonreductive physicalism. It is a variant of Kim's argument that makes explicit an assumption of a causal theory of properties. We have tried to show that Sydney Shoemaker's recent attempt to harmonize the two positions fails. We believe that such failure was inevitable, given Shoemaker's unusual combination of physicalist commitments, a causal theory of properties, and an abundant, rather than sparse, ontology of such properties. It is an unstable compound. The causal theory requires that properties earn their keep. This appears inevitably to push the philosopher in one of two directions when it comes to macroscopic structures: reduction or elimination, on the one hand (we needn't here adjudicate the claims of these two), or a rejection of one or more of the characteristic claims of physicalism, on the other.

A typical rejection of physicalism includes the denial of the realization, causal completeness, and the supervenience premises. (We'll here set aside the question whether the rejection of physicalism requires rejecting all forms of supervenience, as it is not resolved easily and is not important for our purposes here anyway.[20])

But before we consider what form a rejection of physicalism might take, we must consider an alternative and original proposal defended in a number of places by Carl Gillett. Gillett contends that we can best reject the Kim-style argument against non-reductive physicalism not by rejecting physicalism itself, but by weakening it. He suggests that we may retain realization and supervenience, and reject only completeness.

4. GILLETT'S 'STRONG EMERGENCE'

Gillett dubs his view 'strong emergentism.' On this view, mental properties, like all macro-level properties, are realized microphysically. That is to say, an instance of a mental property is identical to a combination of *other*, micro-physical property instances and relations among them, where the other properties are, in the circumstances, the property's realizers.[21] Strong emergence occurs where *microphysical* properties contribute *different* fundamental causal powers to their microphysical individuals precisely when these properties realize certain properties.

A schematic example: microphysical property **L** confers upon the microphysical entity that bears it only powers α, β, and γ in all circumstances *except* when it

[20] See the discussion in O'Connor and Wong (2005).

[21] See Gillett (2006b: 275, 280, 281, 282). See also Gillett (2002) for the proposal and defense of his preferred account of realization.

realizes property **H**, in which case it confers power δ. Now, notice that it's still the case that only microphysical properties contribute the fundamental, irreducible causal powers. (We'll hereafter let the qualifiers 'fundamental' and 'irreducible' be implicit.) But Gillett argues that the realized property **H** is nevertheless causally *efficacious*, for three reasons:

i) **H** *non-causally determines* **L** to contribute the δ power to its bearer,

ii) **H** is a necessary member of a set of factors jointly sufficient for the contribution of δ to an individual, and

iii) positing **H** as a causally efficacious property is necessary if we are to 'account for' the relevant microphysical individual's having power δ.[22]

Furthermore, since in cases of strong emergence, we have realized properties, such as **H**, determining the contribution of causal powers (like δ) by their micro-level realizer properties (**L**), these count for Gillett as cases of 'downward determination' by **H**.[23]

Such, in brief, is Gillett's suggested picture of 'strong' emergence. How does it fare as an explication of robust nonreductionism, and is it preferable to non-realizationist varieties of emergence? Notice first that since **H** is a realized property, the microphysical property **L** whose activity it non-causally determines is part of **H** itself. This may seem to result in an objectionable circle of determination relations, but it does not. While an instance of **L** contributes to the constitutive determination of an instance of **H**, the latter is not thought to similarly bear some kind of ontological priority to the former. Instead, **H** determines which causal powers **L** shall confer in the context.

Our basic criticism is this: Gillett's strong emergence provides at best a very attenuated form of causal efficacy for mental properties. They do not produce (non-derivatively) any event or even trigger some other causal power into activity. They seem to be simply the occasion on which microphysical properties act in unusual ways (i.e., ways departing from their nearly ubiquitous manner of activity). In fact, from the perspective of a causal powers theorist, **H** in our example seems but a handy name for the sort of circumstances in which **L** confers δ; it answers one sort of 'when' question. But it's hard to see what's gained in explanation by insisting on accepting an emergent realized property into our ontology. In fact, this insistence plausibly obscures, for (when combined with 'determination' talk) it makes it look as if there's some light being shed on *how* and *why* **L** confers δ when there's not. The sober metaphysical truth seems to be that whenever **L** is co-instanced with certain other properties in a certain way—where that way and those other properties can be wholly specified in microphysical terms—then **L** confers δ. We can introduce the term '**H**' as a label for this type of scenario. But in doing so, we wouldn't be accounting for anything that we

[22] Gillett (2003: 102); Gillett (2006b: 274, 281, 282, 287).

[23] Gillett (2003a).

hadn't previously accounted for in speaking only of microphysical properties and relations, and we wouldn't have gotten one step closer to understanding how or why **L** confers δ in the relevant scenarios. Hence, we don't think we ought to accept Gillett's contention that **H**, understood as an emergent realized property, is a necessary posit in 'accounting for' the contribution of δ by **L**.[24]

Some will be inclined to reply at this point along the lines of Jerry Fodor's brief on behalf of the standard, non-emergentist variety of nonreductionism: we must recognize **H** as a real, multiply realized, and explanatory property in its own right because otherwise we will fail to capture the commonality of the many different scenarios, microphysically described, in which **L** confers δ. Only here the case for **H** would be bolstered by the fact (*ex hypothesi*) that the *fundamental* dynamics are distinctive in nonreductive scenarios. We are unmoved, but suppose one is inclined to concede the point. Even so, all we would have embraced are mental properties that play a kind of structuring role in the world's dynamics. They do no distinctive causal work—provide no extra causal oomph. There is, indeed, a strong analogy here to the role played by spatial and temporal relations in Newtonian mechanics, as construed by a causal powers theorist.[25] Such relations, one might say, provide a necessary framework for the interplay of dispositional entities, while themselves having no dispositional nature. Surely our nonreductionist physicalist wants more than this by way of the causal relevance of the mental. More than being local, nondispositional constraints on the way fundamental physical causes operate, our beliefs, desires, and intentions themselves directly contribute to the unfolding dynamics of our behavior.

5. A BETTER ACCOUNT OF EMERGENCE

It thus appears that a rejection of the causal completeness tenet of mainstream physicalism will not in itself suffice to secure a robust efficacy for mental properties. We must also reject the realization thesis, and in the context of mental

[24] Gillett buttresses his claims about a non-causal determination relation by comparing the case of emergent properties (like **H**) to what he counts as other cases where there is non-causal determination. Such cases include parts–wholes, realization, constitution, and properties that contribute conditional powers. See Gillett (2003a: 109, 2006a: 6–7, 2006b: 268). Though more deserves to be said in response, we'll here say only that a causal powers theorist is under no obligation to accept, and may have good reason to reject, each of the four additional candidates for non-causal determination relations Gillett proposes. The grounds in favor of this response are, in brief, that (i) questions about which properties count as causally efficacious (in a causal nonreductionist sense) ought to be settled prior to any commitment concerning the first three proposed relations, and (ii) a causal powers theorist inclined toward 'sparseness' with respect to properties will reject the sort of conditional power-conferring properties Gillett invokes (in his 2003a: 101 and 2006b: 279–80, 285–6) as the *relata* of the fourth proposed relation.

[25] Gillett anticipates this analogy by deeming spatial relations entities that (if they exist) 'do not contribute powers themselves' but 'may still determine the contributions of powers to individuals by other properties and relations' (Gillett 2003b: 35).

causation, at least, that is clearly a rejection of physicalism altogether. In our judgment, the best avenue for developing an anti-physicalist view rooted in the rejection of realization and completeness involves a stronger variety of emergence, what is often termed *ontological emergence*.

The term 'emergence' is used to cover a multitude of sympathies (in some cases, sins). So we want to indicate in clear, albeit very abstract, terms what an emergentist picture would look like, in our way of thinking.

Properties are *ontologically emergent* just in case:

(i) They are ontologically basic properties (token-distinct from, and unrealized by, any structural properties of the system).

(ii) As basic properties, they constitute new powers in the systems that have them, powers that non-redundantly contribute to the system's collective causal power, which is otherwise determined by the aggregations of, and relations between, the properties of the system's microphysical parts. Such non-redundant causal power necessarily means a difference even at the microphysical level of the system's unfolding behavior. (This is compatible with the thesis that the laws of particle physics are applicable to such systems. It requires only that such laws be supplemented to account for the interaction of large-scale properties with the properties of small-scale systems.)

In respects (i) and (ii), emergent properties are no less basic ontologically than unit negative charge is taken to be by current physics. However, emergent and microphysical properties differ in that

(iii) emergent properties appear in and only in organized complex systems of an empirically specifiable sort and persist if and only if the system maintains the requisite organized complexity. The sort of complexity at issue can be expected to be insensitive to continuous small-scale dynamical changes at the microphysical level.[26]

We are inclined to further suppose (though this may depend on our inclination to accept a controversial, strong causal explanatory principle) that

(iv) the appearance of emergent properties is *causally originated and sustained* by the joint efficacy of the qualities and relations of some of the system's fundamental parts. (This would involve fundamental properties having latent dispositions to contribute to effects, dispositions that are triggered only in organized complexes of the requisite sort.)

[26] Concepts of emergence have a long history—one need only consider Aristotle's notion of irreducible substantial forms. Their coherence is also a matter of controversy. For an attempt to sort out the different ideas that have carried this label, see O'Connor and Wong (2002). And for a detailed exposition and defense of the notion we rely on in the text, see O'Connor and Wong (2005).

One cannot give uncontroversial examples of emergent properties, of course. Though there are ever so many macroscopic phenomena that seem to be governed by principles of organization highly insensitive to microphysical dynamics, it remains an open question whether such behavior is nonetheless wholly determined, in the final analysis, by ordinary particle dynamics of microphysical structures in and around the system in question.[27] Given the intractable difficulties of trying to compute values for the extremely large number of particles in any medium-sized system (as well as the compounding error of innumerable applications of approximation techniques used even in measuring small-scale systems), it may well forever be impossible in practice to attempt to directly test for the presence or absence of a truly (ontologically) emergent feature in a macroscopic system. Furthermore, it is difficult to try to spell out in any detail the impact of such a property using a realistic (even if hypothetical) example, since plausible candidates (e.g., phase state transitions or superconductivity in solid state physics, protein functionality in biology, animal consciousness) would likely involve the simultaneous emergence of multiple, interacting properties. Suffice it to say that if, for example, a particular protein molecule were to have emergent properties, then the unfolding dynamics of that molecule *at a microscopic level* would diverge in specifiable ways from what an ideal particle physicist (lacking computational and precision limitations) would expect by extrapolating from a complete understanding of the dynamics of small-scale particle systems. The nature and degree of divergence would provide a basis for capturing the distinctive contribution of the emergent features of the molecule.

Now, many contemporary philosophers seem to think that such a view is too extreme to be plausible. When pressed, such critics often cite the alleged consequence that an emergentist view compromises *the unity of nature*. But unity does not require the reductionist vision of the world as merely a vast network binding together local microphysical facts, with a pervasive and uniform causal continuity underlying all complex systems. It is enough that at every juncture introducing some new kind of causally discontinuous behavior, there is a causal source for that discontinuity in the network of dispositions that underlie it. In short: unity in the order of the unfolding natural world need not involve causal continuity of behavior, only continuity of dispositional structure.[28] For the emergentist, the seeds of every emergent property and the behavior it manifests are found within the world's fundamental elements, in the form of latent dispositions awaiting only the right context for manifestation.

[27] For numerous examples of such phenomena, see R. B. Laughlin et al. (2000).

[28] This is not to concede that it is ipso facto a theoretical virtue for a metaphysics that it entails greater unity in nature, nor that it is ipso facto a theoretical vice if the converse is true. The issue of the unity of nature, and the related issue of unity in science, are deep and complex. Our point in the text is that there is *a* kind of unity in nature if the emergentist account we have proposed is correct. For more on the topics of unity in science and nature, see Cat (2007).

We make no assertion one way or the other as to whether anything is like this for any chemical or biological properties, though we note that present evidence allows for the possibility that some perfectly respectable biological and chemical features are ontologically emergent in this way.

We do, however, propose that the conscious intentional and phenomenal aspects of the mind strongly favor an emergentist account. A human person's experiences and other conscious mental states exhibit features quite unlike those of physical objects, whether as revealed in ordinary sense perception or as uncovered in the physical and biological sciences. And the maximally direct nature of our first-person awareness of the intentional and phenomenal features of our conscious states prohibits the a posteriori ascription to them of underlying physical micro-structure hidden to introspection. The upshot of this familiar reflection, if it stands, is that our experiences and other conscious mental states have fundamentally distinctive characteristics. But these very characteristics are also *prima facie* causally efficacious. (Indeed, on a causal powers metaphysics, to countenance them as properties is to accept them as efficacious.) Thus, certain mental properties appear to be (1) resistant to analysis in terms of physical structural properties and so plausibly ontologically basic; (2) causally efficacious; and (3) borne only by highly organized and complex systems. Though we cannot argue the matter at length here, we find extant materialist attempts to overcome this prima facie case to be implausible.[29] (It goes without saying that we take the grounds for an emergentist account of the mental to be defeasible.)

Some philosophers acknowledge that the sort of broadly 'Cartesian' picture sketched above captures how we naively think about conscious experience but contend that it is an illusion. For our part, we think that such philosophers underestimate the difficulties for a theory of empirical knowledge that maintains that we are subject to a radical and pervasive cognitive illusion at the very source of all our empirical evidence. And if the central argument of this paper is correct, then for any of these philosophers likewise committed to a causal powers metaphysic, the seemingly *paradoxical* position of denying the causal efficacy of mental states must be added to those difficulties.

[29] For extended argument on this point, see Timothy O'Connor and Kevin Kimble (manuscript), 'The Argument from Consciousness Revisited.'