

whether dualism is true. Traditional issues of life after death, the existence of freedom and moral responsibility, and the explanatory powers of natural science within its own domain, do not seem to hinge on the answer to the question.

On openness: I think that we do not know enough about the relation between psychology and the natural sciences to take hard lines for or against materialism. I believe that at this point a modest dualism is clearly more reasonable than materialism. But I think that materialistically motivated empirical research is heuristically tenable. The key to any view in this area is openness to empirical exploration and philosophical reflection. Both positions should be developed in a dialectically open spirit.

A modest dualism, however, cleaves more closely to what we know. It does not make warrantless claims. It is more reasonable than compositional materialism, or other sorts of materialism or physicalism, partly through abnegation, and partly through appreciation of the deep differences between rational structures and physical structures. Psychological events have not been shown to have any attributes that are distinctive of physical events. Their primary attributes are not those cited in the natural sciences.

I confess to a more psychological influence on my dualism. I was educated in philosophy in a climate in which materialism had become smug ideological dogma. It dipped with the more supercilious aspects of blind religious orthodoxy. It was not just that herd instinct in philosophy was itself a danger signal. I thought that there was an implicit hypocrisy in the climate. What concerned me was that many philosophers saw any doubt about materialism as *ipso facto* irrational and unscientific. Many still do. Many philosophers exuded a certainty that was out of line with the speculativeness and lack of force in the grounds supporting their positions. Many still do. Such philosophers assume the mantle of science while contravening its letter and spirit. I hope that new generations of philosophers will do better.

## 12

### Descartes' Revenge Part II: The Supervenience Argument Strikes Back

Neal Judisch

#### 1.

For nearly three decades Jaegwon Kim has argued that antireductionist theories of mind are inconsistent with mental causation. Here I aim to show that if Kim's charge against antireductionism is correct his own reductionist view falls prey to the same plight. Specifically, I argue that (i) Kim's theory salvages mental causation only if mental properties are multiply realizable, physically reducible, and have instances that are causally efficacious in virtue of being *mental* property instances, but (ii) his theory is unable jointly to satisfy these conditions. In section 2 I present the Supervenience Argument designed by Kim to prove that psychophysical causation requires reductionism. In section 3 I highlight two principles upon which Kim's charge against nonreductive physicalism essentially relies, and in section 4 I show how a consistent application of these principles places demands upon theories of mental causation that Kim's reductive functionalism cannot meet. The result is that no theory of mind other than type physicalism is consistent with mental causation if Kim's case against antireductionism holds up. If, on the other hand, Kim's reductive functionalism does the trick, antireductive physicalism has nothing to fear from the Supervenience Argument.

#### 2.

The Supervenience Argument incorporates three central assumptions. The first one specifies that the physical world is causally closed:

*Closure.* If a physical event has a cause at *t*, then it has a physical cause at *t*. (Kim 2005: 15)

The second one stipulates that mental properties supervene upon physical properties:

*Supervenience.* If any system  $s$  instantiates a mental property  $M$  at  $t$ , there necessarily exists a physical property  $P$  such that  $s$  instantiates  $P$  at  $t$ , and necessarily anything instantiating  $P$  at any time instantiates  $M$  at that time. (p. 33)

And the third is an exclusion principle expressing the prohibition of systematic overdetermination:

*Exclusion.* If an event  $e$  has a sufficient cause  $c$  at  $t$ , no event at  $t$  distinct from  $c$  can be a cause of  $e$  (unless this is a genuine case of causal overdetermination). (p. 17)

According to Kim, if we further assume that mental properties are neither reducible to nor identifiable with physical properties, what results is a set of propositions inconsistent with the causal relevance of mental properties:

*The problem of mental causation.* Causal efficacy of mental properties is inconsistent with the joint acceptance of the following four claims: (i) physical causal closure, (ii) causal exclusion, (iii) mind-body supervenience, and (iv) mental/physical property dualism—the view that mental properties are irreducible to physical properties. (pp. 21–2)

The reasoning behind this contention is as follows. Suppose we wish to identify a mental property instance,  $M$ , as the cause of a subsequent physical property instance,  $P$ . By *Supervenience* we know that there must be some physical property instance upon which  $M$  supervenes and by *Closure* we know that if  $P$  has a cause at a time  $t$ , it has a physical cause at  $t$ . Let us suppose that  $P$  has a cause at  $t$  and that the physical cause of  $P$  (at  $t$ ) is  $P_0$ , and let us assume that  $P_0$  is the physical property instance upon which  $M$  supervenes.<sup>1</sup> By *Exclusion* we know that  $P$  has no cause other than  $P_0$  unless this is a case of genuine causal overdetermination, which, we will assume, it is not. From this it follows that  $M$  is the cause of  $P$  only if  $M = P$ . But given that no mental property is identical with or reducible to any physical property, it follows that the putative mental cause,  $M$ , is not in reality a cause of  $P$ .<sup>2</sup> Since there is nothing special about  $M$ ,  $P$ , or  $P_0$ , the argument generalizes to show that instances of irreducible mental properties do not have physical effects, so that nonreductive physicalism entails epiphenomenalism: 'That then is the supervenience argument against mental causation, or Descartes's revenge against the physicalists' (Kim 1998: 46).

<sup>1</sup> The assumption is innocuous: if  $M$ 's supervenience base is some physical property instance distinct from  $P_0$ , the closure and exclusion principles still kick in to screen off any causal influence  $M$  might exercise with respect to  $P$ .

<sup>2</sup> It follows, that is, on Kim's supposition that the irreducibility of mental to physical properties entails the nonidentity of mental and physical property instances. More on this just below.

## 3.

I am not interested in disputing the validity of this argument or any of the assumptions it deploys. Instead I want to make explicit two principles Kim relies upon to force the choice between reductionism and epiphenomenalism. After laying them out I will argue that Kim's theory of mental causation is untenable in light of the parameters they set.

It is crucial to keep clear on the distinction between *properties* and *property instances*: Kim himself is quite explicit that properties are 'causal' in virtue of their *instances* being causal:

Properties as such don't enter into causal relations; when we say ' $M$  causes  $M^*$ ', that is short for 'An instance of  $M$  causes an instance of  $M^*$ ' or 'An instantiation of  $M$  causes  $M^*$  to instantiate on that occasion.' (Kim 2005: 39)

And he is just as clear concerning the intended reading of the premiss that 'Mental properties are not reducible to, and are not identical with, physical properties' (p. 34) as it functions within his argument. According to him, just as the claim that ' $M$  causes  $M^*$ ' is short for 'An instance of  $M$  causes an instance of  $M^*$ ', the assumption that ' $M \neq P$ ' only means that this instance of  $M \neq$  this instance of  $P$ ' (p. 42, n. 9). As is well known, however, not all antireductionists endorse this assumption. For these theorists the nonidentity of mental and physical phenomena enters into the picture at the level of *properties*, not at the level of *property instances*. Indeed, Kim's initial formulations of the Supervenience Argument were targeted precisely at such 'token identity' theories. He claimed then, as now, that not just any theory according to which mental and physical property instances are identical suffices to vindicate mental causation, since mental causation requires that mental events produce their effects *in virtue of being the sorts of mental events they are*.<sup>3</sup> But if mental properties cannot be reduced to physical properties any mental property instance identical with a physical property instance is causal only as regards its *physical* characteristics, which is epiphenomenalism near enough. To save mental causation, therefore, what is required is either reduction or identity of mental with physical properties themselves.

Why then does Kim specify that the assumption  $M \neq P$  only means that this instance of  $M \neq$  this instance of  $P$ ? As he realizes, this qualification evidently leaves open a question Kim has tried to close:

Does this mean that a Davidsonian 'token identity' suffices here? The answer is no: the relevant sense in which an instance of  $M =$  an instance of  $P$  requires either property

<sup>3</sup> Thus when we say 'that mental events cause physical events' we intend 'that an event, *in virtue of its mental property*, causes another event to have a certain physical property,' which requires the reduction of mental to physical properties (Kim 1989: 279).

identity  $M = P$  or some form of reductive relationship between them. . . . The fact that properties  $M$  and  $P$  must be implicated in the identity, or nonidentity, of  $M$  and  $P$  instances can be seen from the fact that 'An  $M$ -instance causes a  $P$ -instance' must be understood with the proviso 'in virtue of the former being an instance of  $M$  and the latter an instance of  $P$ '. (Ibid.)

So if a mental property instance is identical to a physical property instance it must be either that the *properties* instantiated are identical or that one of them reduces to the other, otherwise it would not be causal in virtue of its mental features. In the 'relevant sense,' then, no theory according to which mental properties are nonidentical with or irreducible to physical properties can lay claim to the identity of mental with physical property instances either. Thus we have the following principle on *Property instance identification*:

(PII) For any property instances  $x$  and  $y$ , if  $x = y$  then there exist properties  $z$  and  $w$  such that  $x$  is an instance of  $z$  and  $y$  is an instance of  $w$  and either  $z = w$  or there is a reductive relation  $\mathfrak{R}$  such that  $z$  bears  $\mathfrak{R}$  to  $w$ .

What conditions constrain the relationship  $\mathfrak{R}$ ? According to Kim the 'root meaning of reduction' may be summarized as saying that 'if  $X$ s reduce to  $Y$ s, then  $X$ s are nothing over and above the  $Y$ s'. Thus if a property exists but is 'nothing over and above' a physical phenomenon it must be identical with something physical: 'if anything is physically reduced, it must be identical with some physical item' (p. 34). Since  $\mathfrak{R}$  is a relation of physical reduction it follows, in accordance with Kim's stipulation, that any reducible property must conform to this *Reducibility principle*:

(RP) For any  $x$  and  $y$ , if  $x$  bears  $\mathfrak{R}$  to  $y$  then there exists some  $z$  such that  $z$  is physical and  $x = z$ .

A question emerges: Does (RP) imply that (PII) contains a superfluous disjunct? From (PII) it follows that an instance of a property  $M$  is identical to an instance of a property  $P$  only if either  $M = P$  or  $M$  bears  $\mathfrak{R}$  to  $P$ . But if  $M$  bears  $\mathfrak{R}$  to  $P$  it follows from (RP) that  $M =$  some physical item. And what could this 'physical item' be if not  $P$ , given that the properties  $M$  and  $P$  must be implicated in the identity of  $M$  and  $P$  instances? The identification of  $M$  with some other physical item,  $P^*$ , looks to underwrite the identity of  $M$ -instances and  $P^*$ -instances, which is of no obvious use to the identity of  $M$ -instances and  $P$ -instances unless  $P$ -instances are  $P^*$ -instances. But if  $P$ -instances are  $P^*$ -instances then, by (PII), either  $P$  and  $P^*$  are identical, in which case (by transitivity of  $\Rightarrow$ )  $M = P$  after all, or  $P$  and  $P^*$  stand in relation  $\mathfrak{R}$ , in which case a precisely similar series of questions arises.<sup>4</sup>

<sup>4</sup> The question becomes all the more pressing when we consider that Kim's theory of events itself seems to entail the superfluity of the last disjunct in (PII)'s consequent. On that picture any event is an ordered triple consisting of an object, a property and a time, and an event  $(O, x, t) =$  an

The answer must be that there are two nonequivalent ways in which mental and physical property instances may be legitimately identified. The first is via an identification of the relevant properties. But Kim does not accept type physicalism for reasons involving the assumed multiple realizability of the mental within the physical. Still, he believes his approach entitles him to identify mental and physical property instances by way of a reductive relationship between mental and physical properties, and further that this relationship suffices to secure mental causation: 'It may be true—I think it is true—that type physicalism will vindicate mental causation, but it may not be the only position on the mind-body problem that can do this. In my view, functional reduction . . . which, unlike . . . type physicalism, is immune to the notorious multiple realization argument, can also ground mental causation.' (p. 148)

So Kim's (PII) does not demand type-type identities: if a property  $M$  bears  $\mathfrak{R}$  to a property  $P$  we may conclude that any  $M$ -instance is a  $P$ -instance, and this does not require that  $M = P$ . What it does require (in the absence of such identities) is that  $M$  be *functionally reducible* to  $P$ .

I shall have more to say about functional reduction in section 4. For now what requires our attention is the importance of (PII) and (RP) to Kim's strategic posture. Given his insistence that antireductive physicalism entails epiphenomenalism in view of its failure to satisfy (PII), Kim must show how his theory does not run afoul of that principle. But given that type physicalism falls victim to the multiple realization argument he cannot satisfy (PII) by effecting type-type identifications between mental and physical properties. Rather, he must outline a theory of reduction that does not entail such identities, but which nevertheless respects the specification in (RP) that property reduction requires the identification of reduced properties with something physical. To steer a course between type physicalism and antireductionism, therefore, Kim's theory must entail that a mental property  $M$  can be (a) multiply realizable and (b) physically reducible, where (b) demands the existence of a physical item with which  $M$  may be identified. Further, if the theory is to avoid epiphenomenalism it must entail that (c) instances of  $M$  are causally efficacious in virtue of being  $M$ -instances, for the failure to satisfy (c) has been his chief complaint against antireductive physicalism from the start. In what follows I argue that Kim's theory cannot satisfy these three conditions.

## 4.

If Kim can ensure the identity of mental and physical property instances and the antireductionist cannot he must posit as a condition on property instance

event  $(O^*, y, t^*)$  iff  $O = O^*$  and  $x = y$  and  $t = t^*$ . But then there is no hope of identifying mental and physical events by way of the reductive relation  $\mathfrak{R}$  unless it's true that  $M$  bears  $\mathfrak{R}$  to  $P$  only if  $M = P$ , in which case the qualifier concerning  $\mathfrak{R}$  once more appears idle. See Kim (1976).

identification the physical reducibility of mental properties. Moreover, the reduction of a mental property  $M$  demands that  $M$  be identified with some physical item, and the physical item must be such as to imply that  $M$ -instances are causal in virtue of instantiating  $M$ . What could this physical item be? Because  $M$  is multiply realizable it cannot be identified with any singular physical type that realizes  $M$ . But according to Kim, if  $M$  is a functional property a reduction of  $M$  to its physical realizers can be accomplished and this will preserve its causal efficacy. Functional reduction involves three steps. First, we assign the property targeted for reduction a functional definition specifying the typical causal conditions under which the property is exemplified and the typical effects of its instantiation. Second, we search the reduction base domain for properties that implement the causal role definitive of the functionalized property. Third, we construct a theory explaining how the realizers of this property perform the specified causal tasks (pp. 101–2). Then, when a physical property instance  $p$  satisfies the causal profile of a functionalized property  $M$  we may be confident that  $p$  is an instance of  $M$ , in which case the causal powers of this instance of  $M$  will be identical to the causal powers of  $p$ : 'if pain can be functionalized in this sense, its instances will have the causal powers of pain's realizers . . . [and] the problem of mental causation has a simple solution for all pain instances' (pp. 25–6).

Conspicuously absent from this description is any specification of which 'physical item' it is with which the mental property (pain) has been reductively identified. Curiously, while Kim's outline of the procedure begins with the aim of reducing the property pain it moves directly into a discussion of the conditions under which pain is instantiated, which then (in recognition of pain's multiple realizability) leads him to remark that 'Neural bases may differ for different instances of pain, but individual pains must nonetheless reduce to their respective neural/physical realizers,' and the process terminates with the result that 'all pain instances are reduced to the instances of their realizers' (pp. 24–5, italics mine). This is confusing in several respects. To begin, it makes little sense to say that one property instance 'reduces' to another if reduction requires functionalization, since token, non-repeatable events aren't the sorts of things that admit of definition in terms of their 'typical' causes and effects. But if the 'reduction' of an  $M$ -instance to  $P$ -instance simply amounts to their identification, (PII) licenses the identification only if there is 'some form of reductive relationship' between the properties  $M$  and  $P$  themselves. Yet the identification of 'individual pains' with their 'respective realizers' has no tendency to suggest that pain itself has been reduced to anything, still less that it has been successfully identified with 'some physical item' as (RP) demands. Indeed, nothing in the above serves noticeably to demarcate Kim's picture from

classical functionalist construals of the 'token-identity' theory, which he has repeatedly alleged entail epiphenomenalism.<sup>5</sup>

The crucial questions for Kim then are these: have mental properties been eliminated? If not, are they epiphenomenal (are their instances causal, but not in virtue of their mental aspects)? If not—if mental properties may be reductively identified with causally efficacious items—which physical things are they? If either of the first two questions is answered affirmatively Kim's theory can hardly be supposed to ensure that mental events are causal 'in virtue of being mental property instances,' and he gains no ground over the antireductive views he has long opposed. But those two questions can be answered in the negative only if he has a suitable reply to the third: for the conjunction of (RP) and (PII) tells us that unless a mental property is identical with something physical it cannot be reduced, in which case its instances cannot be identified with physical property instances. And from this, according to the Supervenience Argument, epiphenomenalism results.

Kim seems aware that these questions are pressing. Unfortunately, he answers the first two without directly addressing the third:

But what of the causal efficacy of pain itself? . . . The answer is that as a kind pain will be causally heterogeneous . . . Pain, as a kind, will lack the kind of causal/nomological unity we expect of true natural kinds, kinds in terms of which scientific theorizing is conducted. This is what we must expect given that pain is a functional property . . . If the term 'multiple' in 'multiple realizations' means anything, it must mean causal/nomological multiplicity . . . On this reductive account, pain will not be causally impotent or epiphenomenal; it is only that pain is causally heterogeneous. (p. 26)

So according to Kim a mental property  $M$ , upon its functionalization and reduction, is causally heterogeneous and ill-suited to figure into scientific laws, but  $M$  is not thereby eliminated or rendered illegible as a cause. And since the reduction of  $M$  requires its identification with something physical there must be some physical item that matches the profile Kim depicts.

I know of one item that perfectly fits the bill:  $M$  could be identified with the disjunctive property each disjunct of which is a physical realizer of  $M$ . But this evidently is not a suggestion Kim is prepared to endorse:

[I]f we insist on having  $M$  as a disjunctive property, we may end up with a property that is largely useless. What good would it do to keep it as a property when it is not a projectible kind that can figure in laws, and cannot serve in causal explanations? . . . Ex hypothesi,

<sup>5</sup> For example, Robert Van Gulick's antireductive functionalism is almost word for word identical to the 'reductive' theory Kim presents here. See Van Gulick (1992: 172) and compare the discussion of nonreductive functionalism in Kim (1998: 51–6).

[diverse realizers of  $M$ ] are heterogeneous kinds, and if heterogeneity is going to mean anything significant, it must mean causal/nomic heterogeneity. Now... this means that instances of  $M$  are not going to show the kind of causal/nomological homogeneity we expect from a scientific kind. In short, [these properties] are causally and nomologically heterogeneous kinds, and this at bottom is the reason for their inductive unprojectibility and ineligibility as causes. (1998: 109–10)

So disjunctive properties are ruled out. Unfortunately, it seems, everything else is too. For if the heterogeneity and unprojectibility of disjunctive properties is 'at bottom' the reason for their ineligibility as causes (and, indeed, is grounds for their elimination), and if *any* physical item with which  $M$  might be identified will inevitably display these features, there can be no suitable physical candidate for  $M$ 's identification—no candidate that would make  $M$ -instances efficacious in virtue of instantiating  $M$ .

At this point we confront two options: we may eliminate mental properties or we may attempt to retain them. Taking the first option first, suppose we agree with Kim that while there are disjunctive and functional concepts which make possible the application of psychological predicates to physical phenomena, these concepts correspond to no real properties (Kim 1998: 103 ff.). If a mental 'property'  $M$  is identified with such a concept then  $M$  has not been identified with any physical item relevant to mental causation (since 'the problem of mental causation does not concern the causal efficacy of *psychological concepts*' (p. 106)), in which case, as we know from (RP) and (PII),  $M$  has not been reduced and therefore has no causally efficacious instances. But clearly this does not solve the problem of mental causation as Kim wields it against the nonreductivist. For according to him any solution to this problem must entail the nonreductivist. For efficacious in virtue of their mental properties, and *eliminating* mental properties hardly implies that events psychologically described are efficacious in virtue of 'instantiating' them.

So suppose we admit the existence of heterogeneous, non-nomic kinds and identify mental properties with them. This will be of no use either. For even if mental properties were identified with such kinds we cannot by sheer fiat decide that the considerations militating against their causal status can simply be ignored, least of all in this context. As Kim has repeatedly stressed, the competition between irreducibly mental and physical causes is generated by the central role properties occupy within the problem of mental causation, 'for it is in terms of properties and their interrelations that we make sense of... concepts that are crucial in this context, such as law, causality, explanation and dependence' (1989: 270). On his reckoning, any causal transaction must be grounded in a law citing the causally homogeneous physical types those events instantiate. Now if no mental property is identical with a causal/nomic physical property, any physical event instantiating a heterogeneous mental kind  $M$  will also instantiate a causal/nomic physical property  $P$  such that  $M \neq P$ . Clearly, any such event will be causal in virtue of instantiating  $P$ . But even granting this event is both an

$M$ -instance and a  $P$ -instance it does not follow that *its being an instance of  $M$*  is enough to deliver mental causation. Specifically, this maneuver cannot solve the problem of mental causation *as formulated by Kim against the nonreductivist*. So, I claim, neither does it solve the problem as it confronts Kim.

Recall that from Kim's perspective the fundamental problem with 'a Davidson token identity' is its implication that 'the very same network of causal relations would obtain in Davidson's world if you were to... [remove]... mentality entirely from the world' (1989: 269), and it entails this because mental events—despite their identity with physical events—are causal in virtue of the *causal/nomic physical kinds they are* but not in virtue of their mental properties. Yet what is true at 'Davidson's world' is equally true at Kim's if mental properties cannot be identified with causal/nomic physical kinds. Thus, even conceding that Kim can guarantee the identity of mental and physical events this does not vindicate mental causation, because the 'relevant sense' in which mental and physical property instances must be identified *was intended to ensure what nonreductive 'token identity' theories cannot*: that such events are causal in virtue of their mental properties. Of course, it may be that the functionalization of mental properties allows for the identification of their instances with physical events in a principled way. But this does not accomplish the aims for which Kim's psychophysical reduction was originally introduced. What motivated the reductionism was not its potential for delivering property-*instance* identities—*those* identities have never been enough to secure mental causation according to Kim—but rather its promise as a means of resolving the causal competition between distinct mental and physical *properties* and the apparent redundancy of the former as opposed to the latter, so that mental events may be causal *in virtue of* their mental aspects. And Kim's strategy for identifying mental with physical events does nothing to alleviate these difficulties at all. Yet if this is reason enough to deny the comparability of mental causation and antireductionism, it is reason enough to deny its comparability with Kim's reductive functionalism as well.

## 5.

If the Supervenience Argument demonstrates that nonreductive physicalism entails epiphenomenalism it does so only by way of (PII) and (RP). Kim's theory either fails to satisfy these principles or satisfies them in a way that leaves the causal competition between mental and physical properties unresolved. Consistently applied, therefore, the principles constraining the Supervenience-Argument's reasoning lead to the conclusion that type physicalism alone is consistent with mental causation. On the other hand, if Kim judges his theory a success he cannot consistently claim that nonreductive token identity theories entail that 'mentality does no causal work' (p. 269). Either way, Kim's reductive functionalism is no better (and no worse) off than the nonreductive theories he has labored so strenuously against.