Paul Raymont (praymont@ryerson.ca, paulraymont@hotmail.com) Completed: December 5, 2002. Published in *Physicalism and Mental Causation*, ed. Walter and Heckmann (Exeter, UK: Imprint Academic, 2003).

Kim on Overdetermination, Exclusion and Nonreductive Physicalism

Jaegwon Kim argues that nonreductive physicalism entails the causal irrelevance of mental features both to mental and physical effects.[1] My focus will be on Kim's reasons for denying the causal relevance of mental properties to physical phenomena, particularly the behaviour of human bodies. Also, my focus will be on intentional states (e.g. beliefs and desires) rather than qualia.

Suppose that we are trying to explain some bodily movement, for example, why Ed's right arm went up. We answer by referring to his desire to wave to a friend. According to Kim, while this appeal to one of Ed's mental properties seems to account for Ed's behaviour, we should be able to explain the same behaviour solely by appeal to physical features, if any brand of physicalism is true. This is because every good physicalist endorses the principle of closure, according to which it is possible to trace the causal ancestry (if there is one) of each physical event or state without having to leave the physical domain. (Kim 1990, 39; cf. his 1989b, 280 and 1998, 40.) More precisely, "If a physical event has a cause at *t*, it has a physical cause at *t*." (Kim 1993c, 360.) Thus, since the motion of an arm is a physical occurrence, it should be possible to explain it in purely physical terms without having to refer to the (allegedly) nonphysical property of desiring to wave. Of course, there would be no problem if this mental property could be identified with a physical feature after all, but nonreductive physicalists rule this out. They maintain that while the particular instance in Ed of the desire to wave is identical with some particular physical state-token (e.g., a brain state), there can be no reduction of

this, or any other, mental property to a physical property. Thus, we are left with two apparent causes of the motion of Ed's arm, for the mental property that is introduced in our explanation cannot be reduced to any of the physical properties that might be cited in accounting for the same phenomenon. Faced with these distinct explanations, Kim asks which one gives *the real reason* why the effect happened. In his words, "We want to ask: `Which really did it? What's the real story?'" (Kim 1990, 40.) One might question the assumption that there can only be one `real reason' and suggest that our actions might all turn out to be overdetermined. However, Kim argues, even if we are willing on occasion to countenance multiple, overdetermining aetiologies for the same effect, we should surely refrain from implicating mental causation in an implausibly ubiquitous overdetermination, with every one of our actions being a result both of mental and physical factors. (Kim 1989a, 247; 1990, 40; 1998, 44.) Kim concludes that if nonreductive physicalism is true, then mental properties are excluded from being causally relevant to physical effects. In his words, if all human behaviour can be explained by appeal to physical properties, "What causal work is left over for . . . any . . . mental property to do?" (Kim 1993b, 354.)

Kim believes that no intimate connection between mental and physical properties short of identity, not even strong supervenience, will put this problem to rest. Indeed, he takes strong supervenience to be part of the problem (Kim 1993b, 353-7; 1993c, 358-62; 1998, 37-47); for it provides another reason (in addition to the principle of closure) to believe that wherever we find a mental property that is nomologically sufficient for a physical effect, and that thus appears to be causally efficacious with respect to it, there too we shall find a physical property that has an at least equally strong claim to such

efficacy. To see why, suppose a mental property *M* appears to be causally efficacious with respect to the instantiation of a physical property *P**. Given strong supervenience, anything that has *M* must also have a physical property (*P*, *P'*, or *P''*, . . .) from the supervenience base of *M*. Suppose then that the instance of *M* in question is realized by a *P*-instance. Kim infers that if *M* is causally efficacious with respect to *P**, so too is *P*, since the possession of *P* guarantees the "simultaneous" possession of something (viz., *M*) that is nomologically sufficient for *P**, so that *P* is itself nomologically sufficient for *P**. (Kim 1998, 44.)[2] We thus seem to have two properties, *M* and *P*, that stake a claim to being efficacious with respect to the same effect. Assuming that our behaviour is not overdetermined, we must deny the efficacy of one of these properties, and what tips the balance in favour of *P* is the causal closure of the physical realm.

It should be noted that in this criticism of nonreductive physicalism, the denial of overdetermination is a premise that (*together* with the closure principle) helps to yield the conclusion that mental properties are inefficacious. Kim sometimes formulates the principle of closure in such a way as to render this premise superfluous, for while he often interprets the principle of closure to mean only that we *need never* appeal to nonphysical features in explaining a physical effect, he occasionally takes it to entail that we *cannot* ever do so.[3] Clearly, on the latter, stronger reading the principle of closure suffices on its own to preclude the efficacy of nonphysical features. Kim removes this ambiguity in a recent paper, where he endorses the weaker reading of the closure principle. He says, "Physical Causal Closure does not exclude nonphysical causes, or causal explanations, of physical events." (Kim 2001, 276.) So, while he does eventually exclude nonphysical causes, that is only because he denies widespread

overdetermination, and not simply because he accepts the principle of closure. His denial of widespread overdetermination is sometimes presented with his "principle of causal exclusion", according to which, "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)."[4] (*Ibid*.) The possibility of genuine causal overdetermination in the case of human behaviour is then ruled out by the claim that it would be implausibly widespread. Kim's denial of widespread overdetermination is sometimes presented in this way, but not always. For example, in one of his most recent presentations of the exclusion problem (Kim 1998, 37-47), the denial of widespread overdetermination is prominent (Kim 1998, 44-5), but the principle of causal exclusion is only briefly alluded to in an endnote (Kim 1998, 128 n. 23).

Kim sometimes suggests another reason for denying the efficacy of mental properties, for he sometimes argues that even if there were nothing problematic about widespread overdetermination *per se*, any overdetermination involving nonphysical (e.g., mental) properties "may come into conflict with the physical causal closure." (Kim 1998, 45.) His suspicion arises from the fact that the principle of closure is not a mere accidental truth but is instead supposed to have some nomological force, so that it holds in nearby non-actual worlds as well. However, if we allow overdetermination by nonphysical features, then, "A minimal counterfactual supposition . . . can lead to a major change in the world" (*Ibid*.), where the major change is that the physical realm no longer forms a causally closed system. This is because if both *M* and *P* are efficacious with respect to *P**, then *M* will still yield *P** in a nearby possible world that resembles the

actual world in all respects except that in that non-actual world *P**'s occurrence is not overdetermined by *P*.

This is true, but implies no breach of the principle of closure, since *M* strongly supervenes on its physical bases. Given that strong supervenience relations hold across the nomologically possible worlds,[5] anything that has *M* in any such world also has a physical property (other than *P*) from the supervenience base of *M*, and this physical property will (for the reasons set forth in Kim's exclusion argument) have at least as strong a claim to efficacy with respect to *P** as *M* has. We therefore have no contravention of the closure principle. If we really were to consider a world in which *M* is not grounded in a strongly subvening physical feature, we would have to venture beyond the nomologically possible worlds. But then the counterfactual supposition that Kim asks us to indulge is not *minimal* after all, since it directs us to a world that is only logically, but not nomologically, possible. It should not trouble us if the principle of closure fails in such worlds, for surely the modal force of that principle is at best only nomological.[6]

I assume here that any possible world that includes a violation of an actual law of nature is less similar to the actual world than any nomologically possible world, and is thus too dissimilar to the actual world to be the subject of a *minimal* counterfactual supposition. Kim may deny this, and maintain, for example, that our counterfactual supposition will be a minimal one if it points to a world that includes a one-off miracle, wherein a single, physically efficacious instance of *M* occurs without being realized by a physical state, but that resembles the actual world in as many other respects as possible. The concern, then, is that this minimal supposition should yield such a fundamental change, with the physical domain no longer constituting a causally closed system.

However, this is an unsurprising result, for we have simply supposed that a nonphysical cause (*M*) produces a physical effect (*P**), without there being any physical cause that occurred at the same time as *M* and that can account for *P**. That is, we have simply set aside whatever laws dictate the supervenience of the mental on the physical in order to suppose that there is a one-off violation of the principle of closure. Regardless of whether this supposition is minimal or not, it should hardly disturb us that it leads to an abrogation of the principle of closure, since that seems to be exactly the content of our supposition: we just supposed that a gap that appears before the physical terminus of a causal chain is filled only by *M*.

Off Limits or Merely Superfluous?

Kim's exclusion problem is supposed to show that if nonreductive physicalism is true, then we *cannot* legitimately appeal to mental properties when giving causal explanations of physical effects, not merely that we *need not* appeal to them. If he only intended to defend the latter claim, he could have rested his case on the principle of closure, and his repeated denials of overdetermination would have no point. So Kim aims to show that nonreductive physicalism does not simply render mental features superfluous, but actually puts them off limits, in the sense that an ascription of causal relevance to them would violate an important principle (namely, the denial of overdetermination).

This is not always clear in discussions of Kim's exclusion problem. For example, Joseph Levine takes the problem to be that in nonreductive physicalism every mental property turns out to be causally redundant, since it "adds nothing" to the underlying physical state's causal power to produce the same behaviour. (Levine 2001, 26.) For Levine, the lesson of Kim's exclusion problem is that only basic physical properties are

causally efficacious in the strong sense of figuring in the "ultimate ground of all causal transactions" (Levine 2001, 28), while mental properties enjoy a "lesser grade of causal efficacy" (*Ibid*.) that derives from their appearance in lawful regularities (which themselves are offshoots of the more basic laws that cover purely physical states). Thus, for Levine, mental properties *have* some measure of causal efficacy. It's just that their efficacy "adds nothing," in the sense that it makes no novel contribution to the efficacy of the more basic physical properties, since it derives wholly therefrom. Hence mental properties enjoy a merely superfluous efficacy, so that although we need never refer to them in explanations of behaviour, we sometimes *can* do so.

While we may, in the end, find something like this approach to be true, more needs to be said before it can stand as a response to the exclusion problem, for Kim takes the exclusion problem to rule out precisely this sort of position. As he formulates it, the problem is that mental properties are utterly devoid of anything that might plausibly be called *causal* relevance or efficacy, anything strong enough to preserve for them a role in causal explanations. Kim emphasizes this feature of the problem when he criticizes a proposal advanced by Ernest LePore and Barry Loewer. (Kim 1990, 45.) They, like Levine, ascribed to mental properties some weaker grade of causal efficacy (or "causal relevance", to use their preferred locution). (LePore and Loewer 1987.) Against this suggestion, Kim claims that if the physical properties on their own are causally sufficient for the behaviour (as they must be if the closure principle is true), then it is hard to see how the mental properties could be causally relevant *in any sense* to that behaviour. When Kim poses for LePore and Loewer the question, "Is there any leftover causal work for them [mental properties] to do?" (Kim 1990, 45),[7] the answer is supposed to be "no,"

and the implication is that since there is no *leftover* causal work for them to do, they do no causal work at all and thus fail to be causally relevant. For Kim, then, the exclusion problem shows that if the mental does no unique causal work in relation to the physical, so that one *need not* refer to mental features in causally explaining any physical effect, then mental features do *no* causal work in relation to the physical, so that one *cannot* refer to them in causally explaining a physical effect.

This anti-overdetermination stance sits ill with the model of supervenient causation that Kim once held (in Kim 1984a, but only tentatively explored in Kim 1989b, 283, and Kim 1990, 49-53). In a case of supervenient causation, "Macro-event *m* [e.g., a mental event] is a cause or effect of event *e* in virtue of the fact that *m* is supervenient on some micro-event, *n*, which is a cause or effect of *e*."[8] (Kim 1989b, 283.) If one accepts Kim's criticism of LePore and Loewer, then one should ask what leftover causal work there is for *m* to do, given that *n* is already sufficient to produce *e*. In fact, this is precisely what led Kim to abandon supervenient causation, retaining the anti-overdetermination stance at its expense. Thus, he says that while *m* is "not an independent cause of *e*[,] so long as *m* remains a distinct property not identified with *n*, we must . . . still contend with the two purported causes of a single event."[9] (Kim 1993c, 361.) This is why Kim, in *Mind in a Physical World*, introduces the problem of causal exclusion *within* a framework in which mental properties strongly supervene upon physical features. (Kim 1998, 37-47.) He notes that given the causal closure of the physical, any physical effect (that has a sufficient cause) will have a sufficient physical cause. He then asks, "What causal work is left for *m* [the instantiation of a mental property] to contribute?" (Kim 1998, 37.) Again, the implication is that since there is a physical state that suffices for the

effect, there is no further causal work for the mental state to do; and, again, he infers that since there is no further or leftover causal work for it to do, it does no causal work at all. More pointedly, he infers not simply that the mental state does no *further* causal work, but also that it does not even merely duplicate or overdetermine the causal work of the physical state.

This aspect of Kim's argument is troubling, for he assumes that an entity that does no *distinctive* causal work in a given case does *no* causal work in that case.[10] That is, he seems to assume that if one entity is causally sufficient for an effect, then no other entity (dependent or not) can exercise a merely duplicating or overdetermining efficacy with respect to that same effect; and this would rule out there being even a single case of overdetermination. However, occasional cases of overdetermination appear to be at least possible. Indeed, in the concern that he elsewhere voices about overdetermination, Kim himself seems willing to allow for exceptional cases of it and only balks at its being very widespread. (Kim 1989a, 247; 1990, 40; 1998, 44.)

Even widespread overdetermination has not been shown to be problematic. As Ned Block has argued, widespread overdetermination is only worrying when it involves a surprising coincidence of independent causes. (Block 1990, 159; cf. Garrett 1998.) But, in view of the strong supervenience connections among the mental and physical features, there is a network of systematic dependency relations among these states, so that we are confronted with no such coincidence when mental and physical processes culminate in the same behaviour.

Still, there is something strange in the notion that whenever one acts for a reason, there is on hand a multiplicity of entities, individually sufficient causes, each exercising

its efficacy so as to engender the behavioural consequence. I share this intuition, which seems to motivate Kim's rejection of overdetermination. I will not try to undermine this intuition here. Instead, I want to emphasize that unless this aspect of the exclusion problem is addressed, we will have to conclude that mental properties *cannot* be included in an accurate causal story about why the behaviour in question occurred (if mental properties are not identical with physical properties). In the following section, I will focus on Kim's rejection of (widespread) overdetermination and argue that this premise in Kim's argument loses its relevance in the context of a coarse-grained model of events, with the consequence that mental features are not off limits for nonreductive physicalists when explaining physical effects.

What Overdetermination?

In many of his discussions of the exclusion argument, Kim deploys his own model of fine-grained events in which an `event' is a thing's having a given property at a time. These three items (thing, property and time) determine the identity of the event. Thus, for Kim, corresponding to the physical property *P* and the mental property *M* are two events, *p* and *m*, which the two properties respectively help to constitute. Kim adds that although he often speaks of one property causing another, he does so only for the sake of brevity (a practice that has heretofore been followed in this paper), since, strictly speaking, it is events, and not properties, that stand in causal relations to each other. (Kim 1993b, 354 n. 33; 1993c, 361; 1998, 41-2.)

With this framework in place, it should be clear that the behavioural effect, the *P**-instance *p**, is causally overdetermined, since it results from two simultaneous events, *m* and *p*, which individually sufficed to cause it. Thus, in answer to our question why *p**

occurred (or perhaps why it has the property *P**), we receive two explanations, which introduce two *causes* of *p**.

Fair enough, but Kim believes that the exclusion problem, with its attendant worry about overdetermination, arises with equal force even if events are more coarse-grained than he takes them to be. More specifically, he repeatedly claims that the exclusion problem confronts those who adopt Donald Davidson's model, in which events are multi-propertied concrete particulars.[11] (Kim 1990, 39, 41 and 45; 1993a, 25-6; 1998, 33-4.) For example, in a rejoinder to Davidson, Kim worries that if the single event *c* has both the physical property *P* and the mental property *M*, both of which are causally relevant to some effect *e*, then, "*c* turns out to have two properties each of which is causally efficacious in *c*'s causation of *e*."[12] (Kim 1993a, 25.) Kim concludes that proponents of coarse events are faced with the exclusion problem, since there appears to be no causal work left for *M* to do (given the closure principle and the denial of overdetermination).[13] (Kim 1993a, 26.)

It is hard to see why this should be so. Let us suppose that we have asked why *e* occurred. Two answers are offered, one that appeals to *M* and one that appeals to *P*. For a proponent of coarse events, both answers appeal to the same event, *c*, which has both *M* and *P*. We then ask why *c* caused *e*; that is, we ask: "In virtue of which of its properties did *c* produce that effect?" This is how the topic of epiphenomenalism, with its questions about causal relevance or efficacy, is introduced within a metaphysics of coarse events: we ask not simply why *e* occurred, but *why the cause caused* e. Answers to the latter question provide answers to the former question, but the converse does not hold, for explaining why *c* caused *e* is a more demanding way of explaining why *e*

occurred. For example, in some contexts it will be a satisfactory explanation of Jo's falling asleep (i.e., *e*) to say that he took a pill that had the power to induce sleep, for we thereby rule out alternative explanations (e.g., that he had had no sleep the night before)[14] and thus give some new information about the effect's aetiology. However, if we press matters further and inquire as to why taking the pill (i.e., *c*) caused Jo to fall asleep, it is unsatisfactory to reply, following Moliere's charlatan, that the pill put Jo to sleep because it had the power to do so, for such an answer merely recapitulates information that our why-question presupposed. Hence, although the pill's power to put Jo to sleep is a property of the cause that might suitably be adduced when our explanandum is simply *why the effect occurred*, few regard it as a causally relevant or "efficacious" property of the cause, since it cannot be suitably adduced when our explanandum is *why the cause produced that effect*. By contrast, the pill's containing diazepam is a causally relevant property, since it does account for why taking the pill caused Jo to fall asleep. More generally, on a coarse-grained model of events, the causally relevant properties of *c* are those that help us to explain why *c* caused *e* (and not simply why *e* occurred).[15] For properties, then, being "causally relevant" is wholly a matter of possessing a certain type of explanatory relevance.

This is why it is misguided to call causally relevant properties "efficacious" within the context of a coarse-grained model of events, for on that model it is only the bearers of the properties (namely, events), not the properties themselves, that wield any real efficacy, and that may thus be subject to the exclusion principle. I claim not to establish this model here, but only to defend it from Kim's charge that the exclusion problem arises within it. On this model the causally relevant or "efficacious" properties

are just those that are salient in the light of the above-mentioned explanatory project (viz., the project of explaining why the cause caused the effect), and we have no reason to believe that these explanatory items compete with or exclude each other. Thus, we do well to eschew Kim's misleading phrase, "two properties each of which is causally *efficacious* in *c*'s causation of *e*" (Kim 1993a, 25, italics mine) and speak instead of "two properties each of which is *explanatory* of *c*'s causation of *e*."

It is true that we sometimes explain why a cause caused its effect by referring to additional events or states, additional causes, which enabled the cause to do so; for example, we may refer to earlier events that helped to structure present circumstances in such a way that *c* will suffice for the effect (for this variety of explanation see Dretske 1988). But we are not doing that in the case I have described, for, by hypothesis, we are referring only to distinct properties of *c* (*M* and *P*), and not to any other event or state tokens. Neither of these properties of the cause is itself a cause of *e*, and we are certainly not claiming that either of them *caused* the cause to cause *e*. Thus, while we often do introduce new causes when we explain, we are not doing that when we refer to *M* or *P* in explaining why *c* caused *e*. We here offer a `causal explanation' only in the sense that we give an explanation *of*, but not *by means of*, a causal relation, so that we do not multiply causes merely by multiplying such explanations. It is hard to see, then, how our multiple answers in terms of *M* and *P* could imply that there is any causal overdetermination, since these several explanations of why *c* caused *e* introduce no additional causes. Such merely explanatory overdetermination (if it is overdetermination at all) seems benign.

What makes *causal* overdetermination problematic is the notion that several entities are present and duplicating each other's causal work. But no such profligacy arises with merely explanatory overdetermination. We do not face any vexing question about what causal work is done by the items to which we refer in our explanations, for these explanatory items are all properties of the same single cause (or they are facts that involve these properties), the same single "doer" of causal work, and do not themselves purport to be doers of causal work. As a result, none of them can be taken to have already done the causal work, and to have thereby excluded the others from doing the same causal work.

Restricting the relevance of properties to explanatory relevance carries no commitment against robust explanatory realism. We can still require that an appeal to *c* succeeds as a causal explanation only if *c* caused the explanandum event, and only if our explanation introduces a property of *c* that stands in the right relation to the property that figures in the explanandum. The "right relation" is generally thought to involve certain nomological connections or counterfactual dependency relations. Notorious difficulties beset attempts to specify more precisely the nature of these relations. Nevertheless, whatever the "right relation" turns out to be, it is one that obtains between properties independently of—indeed, as a presupposition of—our explanatory practices. As realists, we should acknowledge a network of such relations among properties that exist regardless of whether anyone thinks about them. We can add that when an event causes something, it does so *because* it has a property that bears such a relation to a property of the effect. But this "because" does not indicate a causal relation between a property and its bearer. Hence, even if we adopt this explanatory realism, it remains true that within a

model of coarse events properties (mental or physical) enjoy explanatory relevance, but not real efficacy. Also, the fact that more than one property of the cause stands in the "right relation" for explanatory relevance to the explanandum property is not problematic, especially if the properties of the cause are linked by a strong supervenience relation, so that they are not independent of each other (in which case their bearing the "right relation" to the effect involves no puzzling coincidence).[16]

Earning Their Keep

A critic might accept this shift from causality to explanation, but ask what further *explanatory* work remains for a mental property once we have explained *c*'s causing *e* in physical terms. But in raising this question the critic presupposes that once we have explained something in one way, there is no other way in which to explain it, no alternative means (set perhaps in the conceptual apparatus of another science or discourse) of shedding light on the same explanandum. We have no reason to accept this assumption, especially since the actual world seems transparently to sustain a profligacy of explanatory patterns, the trajectories of several of which often do converge on the same explanandum. The systematic connections (viz., supervenience relations) among the properties in these explanatory patterns show that this merely explanatory overdetermination does not involve any bizarre coincidence; and as long as the overdetermination is *merely* explanatory (i.e., involves no multiplication of cause-event tokens), it does not involve multiple entities duplicating each other's causal work.

Unfortunately, allowing for widespread explanatory overdetermination leaves untouched the following objection. In allowing for the convergence of distinct explanantia upon a single explanandum, we at best only allow for the possibility that

mental properties, *in some cases*, explain things that are also explained by physical features. One might tolerate this, but only on the condition that the overdetermining properties at least *sometimes* explain things that cannot otherwise be explained. That is, one may refuse to countenance the possibility that mental properties, *in every case* in which they explain, explain things that are also accounted for by physical properties. After all (the critic may add), simple parsimony militates against recognizing a stock of properties that *never* do any distinctive explanatory work.[17] In short, mental properties, if they are always explanatorily superfluous, face elimination.

This objection does not rest upon a denial of overdetermination, and thus requires no picture of explanantia as competing against each other for, and threatening to exclude each other from, causal relevance with respect to an explanandum. It is therefore a different threat to the existence of causally relevant mental features from the problem that has been my focus in the foregoing. I would like to outline briefly a strategy for dispelling this fear.

The challenge is to find distinctive explanatory work for mental properties. There are two ways to achieve this aim. We may try to demarcate a range of facts that mental properties alone can explain. Call this the "New Fact" approach. Alternatively, we may say that while there might not be any such distinct explananda, mental properties can nonetheless explain *in a new way* facts that can be explained only in a different way by physical features. I dub this the "New Way" approach. Both strategies begin with the observation that our behaviour has a dual character: it may be considered as mere bodily movement, or as meaningful, intentional action. Thus, an item of behaviour can appear in an explanandum in terms of its purely physical character or as an intentional act. For

example, while my arm's rising is merely the movement of a limb so many centimeters upward, my raising my arm to vote is a richer explanandum.[18] It is in this richness that proponents of both strategies claim to find the distinctive explanatory work that mental properties do.

New Fact proponents (e.g., Marras 1998), in trying to isolate explananda that mental properties alone explain, argue that no wholly physical story will be a satisfactory explanation if we take as our explanandum the behaviour as an intentional action (such as voting); they are compelled by the New Fact model to interpret this to mean that the intentional action is simply not explained, not even causally, by the physical facts. However, while it is true that purely physical accounts are unsatisfactory when we inquire about an intentional action, this does not *entail* the inadequacy of the physical account *as a causal explanation*.[19] For we can characterize the unsatisfactory nature of the physical account by saying that, while it may provide a complete causal explanation of the action explanandum, it does not at the same time rationalize the action. Even a complete causal account may be unsatisfactory if what we wanted was for the action to be explained in this different way; that is, if we wanted it to be *at once* causally explained and rationalized.

This leads to the New Way strategy, where we allow for the possibility that even intentional action explananda are causally explainable in physical terms, but add that they can nevertheless be explained in a different way by mental properties, since it is only in mental terms that they can be causally explained in such a way that they are *thereby* rationalized. In such rationalizing explanation, the explanation does double duty as a justification, in the sense that it shows the action, in the light of the agent's reasons, as a

rational thing to have done. It does not show merely why the action did happen, but why it *ought* to have happened, in approximation to some normative ideal of rationality. Accordingly, some philosophers have compared such explanations to a practical syllogism set wholly in the past tense.[20] (Anscombe 1957; von Wright 1971, chapter 3.)

I do not say that mental properties always, or even usually, perform this function. The point is that they sometimes do so, and that this suffices to distinguish them from properties that cannot rationalize actions, and staves off their elimination.

Of course, some have argued that it is precisely because mental properties can rationalize actions that they cannot also causally explain them. (Anscombe 1957; Dray 1957; Peters 1958; Waismann 1955; Winch 1958.) However, I see no reason why this should be so. Causally relevant properties are explanatory of the cause's production of its effect. As was noted in the previous section, the most popular criteria for determining whether a property is thus causally relevant invoke nomological or counterfactual dependency relations. It seems quite plausible that mental features can pass this sort of test. More generally, whatever test of causal relevance we adopt, it should not be so demanding as to exclude geological, biological, and other higher-level physical properties from being causally relevant; it again seems plausible that any thus accommodating test of causal relevance is one that mental properties can pass. I should emphasize that I am not endorsing a particular test of causal relevance and arguing that mental features pass that test. In this respect my method resembles Kim's: he does not proceed by first endorsing a test of causal relevance and then arguing that mental properties fail to meet *that* test; instead, his exclusion problem is supposed to be that *whatever* (plausible) test we adopt, it will be satisfied by physical properties, which then threaten to exclude

mental features from being causally relevant. Against this I maintain that any reasonable test of causal relevance is one that mental features sometimes meet, and that the mere fact that physical properties also meet that standard of relevance to the same explananda impugns neither the relevance nor the distinct status of the mental properties.

With mental features on hand as a distinct stock of properties, it will be an empirical question in a given case whether a mental feature meets the standard for being causally relevant to a given explanandum. It is possible for mental features to do so *without* at the same time rationalizing the explanandum in question.[21] More specifically, they may be causally relevant to some merely physical explananda, facts that do not involve the instantiation of intentional action properties and are therefore not susceptible of being rationalized. I take this to be a virtue of this (admittedly programmatic) approach. For example, if I wave my arm, this action involves my arm's rising, which is a mere bodily motion. This latter is a physical explanandum, and yet it is one that my desire to wave should be capable of causally explaining. (At least, it should be if we take a fairly determinable physical explanandum such as my arm's rising, and not, say, my arm's rising exactly seven-and-a-half centimeters.) Also, there are all sorts of physical facts that obtain in the wake of my actions and that seem to be causally explainable by appeal to the reasons for those actions. For instance, after I have poured some milk, the milk carton will occupy a different position in the refrigerator, its contents will be diminished, and so on. It seems perfectly reasonable to allow that the mental features that led me to pour the milk are causally relevant to these mundane physical facts. More generally, it is an important part of our self-image that our agency leaves its mark on the world, and this remains the case even when the world is considered simply in physical

terms. Hence it is an important virtue of an account if it allows for mental properties to be causally relevant even to physical explananda.[22]

When they are causally relevant to merely physical explananda, mental properties are strictly redundant, making no novel explanatory contribution as compared with the merely physical explanantia. I hope to have shown that there is nothing problematic about such explanatory overdetermination, as long as the overdetermining properties *sometimes* make a novel explanatory contribution. This last condition is met, I suggest, because of the other cases in which mental properties do novel explanatory work by rationalizing intentional actions.

Given the principle of closure, the merely physical explananda to which mental features are relevant will also be causally explainable in wholly physical terms. This may lead one to complain that if earlier physical facts can account for all later physical facts (e.g., all movements of human bodies), then there is no need to appeal to mental features in explaining any physical facts. What explanatory work is left for the mental to do? However, this is, at best, merely a point about the pragmatics of explanation. Having uncovered the full physical story behind our bodies' behaviour, we may feel no need to pose any further "why"-questions about it, and thus no need to appeal to mental features in answering such questions. But that will not cause the nomological or counterfactual connections between mental and behavioural types to dissolve, for surely those are mind-independent connections. More generally, if we adhere to realism about causal relevance (as described at the end of the previous section), then whether that explanatory connection obtains between types should not depend on whether we find it useful or enlightening to consider it or refer to it in our speech acts of explaining.

Conclusion

Kim has not shown that there is anything problematic in the notion of multiple explanantia bearing down on a single explanandum—at least, not in the context of a coarse-grained model of events. Merely to allow for such explanatory overdetermination is not yet to vindicate the causal relevance of mental properties, for it must also be shown that they, at least some of the time, make a novel explanatory contribution. This contribution is best described in terms of how, and not in terms of what, they explain. Once they have in this way been established as a distinct stock of properties, there is no reason to deny that they stand in causal relevance relations to both mental and physical explananda.[23]

Notes

[1] In my presentation of Kim's argument I rely chiefly on Kim 1989a; 1989b; 1990; 1992; 1993a; 1993b; 1993c; and 1998. 1989a, 1989b, 1992 and 1993b are reprinted in Kim 1993d. Page references to these three papers are to the 1993d volume.

[2] If I understand Kim correctly, his point is that since the instantiations of *P* and *M* are simultaneous, it is better to regard them as being related to each other by a realization relation, not by causation; so *P* is a nomologically sufficient condition for *P** with just the same proximity to it in the causal chain as is enjoyed by *M*; hence *P* has at least as strong a claim to being efficacious with respect to *P** as *M* does. Cf. Kim 1993b, 354.

[3] The weaker version of the principle of closure is presented when Kim says, "If a physical event has a cause at *t*, it has a physical cause at *t*." (Kim 1993c, 360.) Consider also the following: "If we trace the causal ancestry of a physical event, we need never go outside the physical domain." (Kim 1989b, 280.) Cf. Kim 1990, 39 and 45. He expresses the stronger closure principle when he says, "No causal chain will ever cross the boundary between the physical and nonphysical." (Kim 1998, 40.) Kim's alternation of these two versions of the closure principle is described in Putnam 1999, 215 n. 17; cf. Clarke 1999: 312. E. J. Lowe documents a similar vacillation by other authors in his 2000: 573-4. Cf. Montero 2003. I have confined myself to distinguishing among versions of the closure principle solely on the basis of whether it is taken to rule out a mental item's figuring in the causal history of a physical effect. Sven Walter has pointed out to me that Kim's denial that causal chains ever cross the boundary between the physical and nonphysical implies an even stronger principle, one that rules out a physical cause's production of a mental effect.

[4] This version of the principle differs from the version of it that appears in Kim's earlier work. According to the earlier version, "No event can be given more than one *complete* and *independent* explanation." (Kim 1989a, 239, italics in the original.) Again, "A cause, or causal explanation, of an event, when it is regarded as a full, sufficient cause or explanation, appears to *exclude* other *independent* purported causes or causal explanations of it." (Kim 1989b, 281, italics in the original.) The importance of the substitution of "distinct" for "independent" in the more recent version of the exclusion principle is clear when we consider Kim's statement that supervenient properties are *distinct* from the properties on which they supervene but are not *independent* of them (Kim 1989a, 251). In view of this, only the more recent version of the exclusion principle precludes the efficacy of the supervening mental properties. In fact, the adoption of this

more recent version of the exclusion principle is what led Kim to abandon his earlier model of supervenient causation (Kim 1993c, 361).

[5] Strong supervenience relations hold at least across the nomologically possible worlds. Kim says that the modality at issue in strong supervenience is either nomological, metaphysical, or logical, but does not further specify it (Kim 1998, 10; cf. Kim 1998, 39). However, in allowing for worlds in which *M* occurs without being grounded in a physical property, he seems to rule out the stronger of the modalities (at least in the case of the strong supervenience that links mental to physical properties).

[6] This objection was presented in an earlier version of this paper at the 2000 meeting of the Western Canadian Philosophical Association, where I received helpful commentary from Bob Bright and Jeff Foss (commentary which gave rise to the following paragraph). Similar criticisms of Kim can be found in Crisp and Warfield 2001: 314; and in Loewer 2001: 319-20.

[7] Kim poses similar questions in his 1993a, 26; 1993b, 354-5; 1993c, 361; and 1998, 37.

[8] I have used "*e*" where Kim uses "*E*" in this quotation. Kim says that while it is really *instances* of properties (i.e., events) that are causes and effects, he sometimes, for the sake of brevity, speaks of properties as the causal relata. (Kim 1993b, 354 n. 33; 1993c, 361; 1998, 41-2.) I think this leads him on occasion to use capital letters (usually reserved for properties) where lower-case letters would make his meaning clearer. (The lower-case "*m*" and "*n*" have been preserved, unaltered from the quoted text.)

[9] I have changed the lettering in this passage.

[10] In this connection, Ausonio Marras says, "But dependent, or `inherited', causal efficacy is causal efficacy nonetheless: why suppose that it is merely `apparent' or `epiphenomenal'?" (Marras 2000, 148.) My criticism of Kim is more general. To wit, my concern is that Kim's reasons for denying the efficacy of dependent causes threaten to prove too much, for they would, if sound, establish the inefficacy of *all* apparently overdetermining causes (dependent or not), thereby rendering all overdetermination impossible.

[11] For Davidson's model of events, see Davidson 1967a and 1967b.

[12] The lettering in this quotation has been changed. Kim uses "*m*" instead of "*c*" for the event that has both *M* and *P*.

[13] In a more recent passage, Kim adopts Davidson's model and speaks of the same event (*m*) as falling under both the mental and physical kinds, *M* and *N* (Kim 1998, 33-4). In a footnote to that passage, Kim maintains that even on this model one still confronts the exclusion problem, since, "No causal work is left for *M* [the mental property]." (Kim 1998, 126 n. 6.)

[14] Following here Ned Block's suggestion in Block 1990, 163.

[15] I develop this line of thought more fully in Raymont 1999.

[16] One may question whether the explanandum property counterfactually depends upon the cause's mental feature. However, if the cause had lacked its mental feature, it would also have lacked the physical property on which it supervenes (since the subvening physical property is sufficient for the supervening property); in which case, the explanandum property would not have been instantiated, assuming that the subvening physical feature really is causally relevant to it. Granted, it seems that if the subvening physical feature were present in the absence of the supervening mental feature, the explanandum property would still have been instantiated. But, given that the subvening feature is held to be at least *nomologically* sufficient for the supervening property, this last bit of counterfactual supposing takes us beyond the nomologically possible worlds; hence it is of dubious relevance to questions of causal-explanatory relevance in the actual world. Cf. LePore and Loewer 1987 and 1989.

[17] I think this suspicion of properties that are always explanatorily superfluous is one of the concerns at issue in Kim's discussions of the Causal Inheritance Principle. (1992, 326-7; 1993b, 355; 1993c, 362-3; Kim 1998, 54-6.)

[18] The example is borrowed from Brodbeck 1963: 310.

[19] I allow that physical properties might causally account for intentional actions only because I think it likely that the former will meet any plausible standard for causally explaining the latter, and not because of the principle of closure. The closure principle requires only that each *physical* explanandum (viz., each case of an event's having a physical property) be causally explainable in physical terms. While intentional action properties certainly depend upon physical features for their realization (e.g., an arm-raising depends on an arm's rising), it does not follow that they are themselves physical properties. Hence an event's possessing an intentional action property need not be regarded as a physical explanandum. Cf. Marras 1998.

[20] Kim nicely elucidates this sort of explanation in Kim 1984b, but sees it as a "mode of understanding actions which is not *causal-predictive*." (Kim 1984b: 316, italics in the original.)
[21] This simply results from the fact that physical properties can meet the standard for causal relevance, but do not rationalize explananda; therefore, the standard for causal relevance cannot have built into it any requirement of rationalization.
[22] New Fact proponents typically rule this out. E.g., Marras (1998) seems to endorse a kind of parallelism, in which mental properties are causally relevant only to intentional action explananda, while physical properties are relevant only to physical explananda. Cf. Thomasson 1998: 191-2.
[23] Thanks to Brian Garrett, Danny Goldstick, Bill Seager, Sonia Sedivy, Sven Walter, and especially Ausonio Marras for comments and conversations that contributed to this paper. I would also like to thank Jaegwon Kim, who was the commentator when I presented parts of this paper at the Eastern Division meeting of the American Philosophical Association in 2001.

References

Anscombe, G. E. M. (1957). *Intention*. Oxford: Basil Blackwell.

Block, Ned. (1990). Can the Mind Change the World? In George Boolos (ed), *Meaning and Method: Essays in Honor of Hilary Putnam*. Cambridge: Cambridge University Press.

Brodbeck, May. (1963). Meaning and Action. *Philosophy of Science*, **30**, 309-24.

Clarke, R. (1999). Nonreductive Physicalism and the Causal Powers of the Mental. *Erkenntnis*, **51**, 295-322.

Crisp, Thomas M. and Warfield, Ted A. (2001). Kim's Master Argument (Review of Kim's *Mind in a Physical World*). *Nous*, **35**, 304-16.

Davidson, Donald. (1967a). The Logical Form of Action Sentences. In Davidson 1980. First published in Nicholas Rescher (ed), *The Logic of Decision and Action*. Pittsburgh: University of Pittsburgh Press.

Davidson, Donald. (1967b). Causal Relations. In Davidson 1980. First published in the *Journal of Philosophy*, **64**, 691-703.

Davidson, Donald. (1980). *Essays on Actions and Events*. Oxford: Clarendon Press.

Dray, William. (1957). *Laws and Explanation in History*. London: Oxford University Press.

Dretske, Fred. (1988). *Explaining Behavior*. Cambridge, MA: The MIT Press.

Garrett, Brian Jonathan. (1998). Pluralism, Causation and Overdetermination. *Synthese*, **116**, 355-78.

Kim, Jaegwon. (1984a). Epiphenomenal and Supervenient Causation. In Kim 1993d. First published in *Midwest Studies in Philosophy*, **9**, 257-70.

Kim, Jaegwon. (1984b). Self-Understanding and Rationalizing Explanations. *Philosophia Naturalis*, **21**, 309-20.

Kim, Jaegwon. (1989a). Mechanism, Purpose, and Explanatory Exclusion. In Kim 1993d. First published in James E. Tomberlin (ed), *Philosophical Perspectives 3, Philosophy of Mind and Action Theory*. Atascadero, CA: Ridgeview Publishing Co., 1989.

Kim, Jaegwon. (1989b). The Myth of Nonreductive Materialism. In Kim 1993d. First published in *Proceedings of the American Philosophical Association*, **63**, (1989): 31-47.

Kim, Jaegwon. (1990). Explanatory Exclusion and the Problem of Mental Causation. In Enrique Villanueva (ed), *Information, Semantics and Epistemology*. Oxford: Basil Blackwell.

Kim, Jaegwon. (1992). Multiple Realization and the Metaphysics of Reduction. In Kim 1993d. First published in *Philosophy and Phenomenological Research*, **52**, (1992): 1-26.

Kim, Jaegwon. (1993a). Can Supervenience and `Non-Strict Laws' Save Anomalous Monism? In John Heil and Alfred Mele (eds), *Mental Causation*. Oxford: Clarendon Press.

Kim, Jaegwon. (1993b). The Nonreductivist's Troubles With Mental Causation. In Kim 1993d. First published in John Heil and Alfred Mele (eds), *Mental Causation*. Oxford: Clarendon Press, 1993.

Kim, Jaegwon. (1993c). Postscripts on Mental Causation. In Kim 1993d.

Kim, Jaegwon. (1993d). *Supervenience and Mind*. Cambridge: Cambridge University Press.

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

Kim, Jaegwon. (2001). Mental Causation and Consciousness: the Two Mind-Body Problems for the Physicalist. In Carl Gillett and Barry Loewer (eds), *Physicalism and Its Discontents*. Cambridge: Cambridge University Press.

LePore, Ernest and Loewer, Barry. (1987). Mind Matters. *Journal of Philosophy*, **84**, 630-42.

LePore, Ernest and Loewer, Barry. (1989). More on Making Mind Matter. *Philosophical Topics*, **18**, 175-191.

Levine, Joseph. (2001). *Purple Haze: the Puzzle of Consciousness*. New York: Oxford University Press.

Loewer, Barry. (2001). Review of Kim's *Mind in a Physical World*. *Journal of Philosophy*, **98**, 315-24.

Lowe, E. J. (2000). Causal Closure Principles and Emergentism. *Philosophy*, **75**, 571-85.

Marras, Ausonio. (1998). Kim's Principle of Explanatory Exclusion. *Australasian Journal of Philosophy*, **76**, 439-51.

Marras, Ausonio. (2000). Critical Notice of Kim's *Mind in a Physical World*. *Canadian Journal of Philosophy*, **30**, 137-60.

Montero, Barbara. (2003). Varieties of Causal Closure. This volume.

Peters, R. S. (1958). *The Concept of Motivation*. London: Routledge and Kegan Paul.

Putnam, Hilary. (1999). *The Threefold Cord: Mind, Body, and World*. New York: Columbia University Press.

Raymont, Paul. (1999). *An Idle Threat: Epiphenomenalism Exposed*. Ph. D. Diss. Toronto: University of Toronto.

Thomasson, Amie. (1998). A Nonreductivist Solution to Mental Causation. *Philosophical Studies*, **89**, 181-91.

Von Wright, Georg Henrik. (1971). *Explanation and Understanding*. Ithaca, NY: Cornell University Press.

Waismann, Friedrich. (1955). Language Strata. In A. G. N. Flew (ed), *Logic and Language* (Second Series). Oxford: Basil Blackwell. Originally presented to the Jowett Society at Oxford in 1946.

Winch, Peter. (1958). *The Idea of a Social Science and Its Relation to Philosophy*. London: Routledge and Kegan Paul.