Some Problems for Reductive Physicalism

Pierre Jacob
Institut Jean Nicod,
UMR 8129
CNRS/EHESS/ENS
Paris, France

Jaegwon Kim's book, Mind in a Physical World, is a dense, sustained and powerful critique of nonreductive physicalism. Nonreductive physicalists are monist physicalists. Hence, they reject Cartesian substance dualism. But since they reject the identity between mental and physical properties, they are property dualists (or pluralists). The nonreductive physicalists' key word here is 'autonomy': according to all versions of nonreductive physicalism, the mental is autonomous. This claim, however, is not unambiguous: it may mean either that mental properties make a causal difference or that they are not nomic.

Davidson's anomalous monism, for example, is a version of nonreductive physicalism. It does not purport to meet the demands of mental realism. It relies instead on mental anomalism, i.e., the view that there can be neither strict psychological nor strict psychophysical laws: mental predicates cannot appear in genuine nomic correlations. Hence, the autonomy of the mental consists in its being anomalous. Davidson's main reason for embracing mental anomalism is the normativity of the mental: unlike the application of physical predicates, the application of mental predicates is governed by normative constraints such as rationality. On this view, the

autonomy of the mental precludes a science of the mental or at the very least it points towards the limits of a scientific understanding of mental phenomena.

According to orthodox functionalism, mental properties are higher-order physical properties which can be multiply realized by lower-order physical properties. A higher-order functional property in general is defined as the possession of a first-order physical property meeting a set of causal conditions. If a mental property is "functionalizable", then its instantiation can both be the effect of either some sensory input or the instance of some other mental property and the cause of either the instance of some other mental property or of some behavioral output. According to Putnam's celebrated multiple realization argument, since a higher-order physical property can be realized by a disjunction of distinct lower-order physical properties on different occasions, it is not identical with any lower-order physical property which happens to realize it (or on which it supervenes) on a given occasion. So functionalism too is a version of nonreductive physicalism. Since orthodox functionalists are mental realists, they are committed to the causal efficacy of mental properties. Thus functionalism secures the autonomy of mental properties: they can be both functional and nomic or projectible because psychological causal generalizations involve mental kinds, not their disjunctive physical realizers. Hence, the autonomy of the mental is the autonomy of scientific psychological explanations.

No doubt at the beginning of the book Kim accepts the motivations of mental realism (e.g., p. 31). This is why he rejects Davidson's anomalous monism and why on his view, it is a condition of adequacy on any version of physicalism that mental properties must turn out to be causally efficacious. Kim's goal in the book is to show that the orthodox functionalists cannot have it both ways: mental properties cannot both be irreducible (or autonomous) and causally efficacious. Importantly, Kim's argument is directed against anti-reductionism, <u>not</u> against the functionalist conception of mental properties.

Unlike orthodox functionalists, Kim does not think that if a property can be functionalized, then anti-reductionism is vindicated. On the contrary, he thinks that

functionalism paves the way for the identification of the functional property with the physical property which realizes it on a given occasion. He thinks so because if functionalism is true of e.g., property F*, then there exists a pattern of causal/nomic relations whereby instances of F* are produced when and only when some other property F is instantiated and which in turn causes instances of property F**. The reason why functionalism entails reductionism is that once such a pattern of causal/nomic relations has been established, it becomes possible to find underlying physical mechanisms which occupy the various causal roles.

If so, then the functional property enjoys the very causal powers of the physical property with which it is identical and mental realism is saved by reductionism. The principle which encapsulates Kim's reductionist interpretation of functionalism is his "causal inheritance principle" which denies the explanatory autonomy of mental properties (p. 54) and according to which a second-order functional property can have no more causal powers on a given occasion than its physical realizer on this occasion. If a mental property is not functionalizable, then (on pain of dualism) the choice is between eliminativism and epiphenomenalism. Kim calls "emergentist" the anti-reductionist's claim that a given property is intrinsic and not open to a functional construal. Kim's own suspicion is that, unlike intentional (or content) properties, phenomenal properties (or qualia) are indeed intrinsic properties which will resist functionalization and therefore reductive physicalism.

Kim's argument against nonreductive physicalism proceeds in two complementary stages. His first and main argument is "the supervenience argument". I will examine it in some details. Then, I'll make some brief comments on the second argument which can be called "the multiple realization argument revisited". And finally, I'll raise a question about Kim's fundamental attitude with respect to mental realism.

The purpose of the supervenience argument is to show that nonreductive physicalism is doomed by the problem of causal exclusion. The argument has the

general form of a dilemma: either mind-body supervenience holds or it does not.[1] Kim agrees with nonreductive physicalists that physicalism would not survive the failure of supervenience. I agree with him. So let's assume that mind-body supervenience holds.

Consider a case of mental-to-mental causation: suppose that the instantiation of mental property M causes the instantiation of another mental property M*. Given the assumption of supervenience, M* supervenes on some physical property P* which realizes it. And now Kim asks: what prompted the instantiation of M*? Given the contribution of P*, what was the contribution of M to the instantiation of M*? How can the instance of M and the instance of P* collaborate in the process whereby M* was exemplified? The only sensible answer open to the nonreductive physicalist seems to be: M caused the instantiation of M* by causing the instantiation of P*. Hence, the nonreductive physicalist is committed to assuming that cases of mental-to-mental causation presuppose what Kim called "downward causation", i.e., the possibility of mental-to-physical causation.[2]

I agree with Kim that answering the question raised in the above paragraph is a challenging task for the nonreductive physicalist. However, there is, I think, a plausible answer available to the nonreductive physicalist — an answer which derives its inspiration chiefly from Kim's own critique of Searle's "biological naturalism" (pp. 47-50). The nonreductive physicalist ought to deny that the relation between the instance of P* and the instance of M* is causal. As Kim puts it (p. 44), "in general, the relation between base properties and supervenient properties is not happily construed as causal. For one thing, the instantiations of the related properties are simultaneous, whereas causes are standardly thought to precede their effects; second, it is difficult, perhaps incoherent, to imagine a causal chain with intermediate links, between the subvenient and the supervenient properties". Assuming that M* strongly supervenes on P*, not

---

[1] As Kim's definition makes clear, the relevant notion is strong supervenience since it involves the claim that "necessarily anything with (subvenient property) P at a time has (supervenient property) M at that time" (p. 39).

[2] "The Nonreductivist's Troubles with Mental Causation", in Kim, Supervenience and Mind, Cambridge: Cambridge University Press, 1993.

only is the instantiation of the former simultaneous with the instantiation of the latter, but anything which exemplifies the latter at t must necessarily instantiate the former at t. There is no possible world in which the latter is exemplified at t and the former is not. Kim, I think, was wrong when he wrote in an ancestor paper to chapter 2 of this book that "the situation is essentially identical to the situation we face when we are given two distinct independent causes for one and the same event, each claimed to be a sufficient cause".[3] It follows that the nonreductive physicalist must presumably recognize that the alleged causal relation between M and M* must be factorized into two sub-relations only one of which is causal: the instance of M causes P* which in turn realizes M*. So the causal relation between M and P* does not compete with the causal relation between P* and M* for there is no causal relation between P* and M*.

This is not to say that the nonreductive physicalist is out of the wood. For we now reach the second step of the argument which involves the principle of the causal closure of the physical: any instance of a physical property must have a physical cause. Given the assumption of mind-body supervenience, mental property M itself supervenes on some physical property P which realizes it. So the question is: given that an instance of P causes the instantiation of P*, what causal work is left for M? Unless he is willing to violate the principle of the causal closure of the physical, the nonreductive physicalist seems bound to grant that property M is epiphenomenal in the process whereby P* gets instantiated. The conclusion of the supervenience argument is that if physicalists want to save the causal efficacy of mental properties, they should opt for reductionism.[4]

I learnt, I think, two important things from pondering over Kim's supervenience argument and his reactions (in chapter 3) to some of the responses offered on behalf of nonreductive physicalism.

---

[3] Ibid., footnote 30, p. 351.

[4] My earlier comment on the first stage of the argument corroborates, I think, Kim's conclusion: the relation between the instance of M* and the instance of P* is not causal and necessarily if P* is instantiated then so is M*. So it follows that the nomic connection between instances of P* and instances of M* is not causal. Rather it must be some kind of identity such as "water = H_2O".

First of all, I agree with Kim that the problem of causal exclusion is about causation, not about explanation. The issue is: how do different kinds of properties relate to one another? Any attempt to overcome this problem by distinguishing different kinds of explanation will fail to engage with the metaphysical issue generated by the metaphysical assumption that mental properties are irreducible.[5] This, I think, applies across the board: Jackson and Pettit's distinction between process and program explanations is explicitly designed to vindicate the "relevance", not the causal efficacy, of higher-order physical properties — and relevance is an epistemic notion. To argue (as Block and Antony and Levine have) for the causal powers of a higher-order chemical property such as dormitivity from its involvement in a generalization such as "dormitive substances provoke traffic accidents" is, I think, to switch from causation to explanation. Of course, to learn that e.g., antihistamines and tricylics have something in common is to learn something useful. However, if it causes a particular traffic accident, the ingestion of a substance could only do so in virtue of its chemical property.[6]

Second of all, I think that Kim's response to "the generalization argument" is quite effective. If correct, this argument would undermine "the spirit of causal exclusion reasoning".[7] It relies on the following conditional: if mental properties are deprived of causal efficacy by their physical realizers, then all macrolevel special science (chemical, geological or biological) properties should be similarly deprived of causal efficacy by fundamental microlevel physical properties. Since nobody in his right mind would accept the consequent, the antecedent can be safely rejected. As Kim correctly points out, the conditional confuses the distinction between levels (of reality) and the distinction between orders (of properties). Macrolevel special scientific properties belong to entities different from the entities which exemplify fundamental microlevel physical properties. This is why the problem of causal exclusion does not apply: the causal efficacy of the former does not compete with the causal efficacy of the latter.

---

[5] I confess that I earlier missed the point.

[6] Cf. N. Block, "Anti-Reductionism Slaps Back", in J.E. Tomberlin (ed.) Philosophical Perspectives, 11, p. 174, Oxford: Blackwell, 1997, and L. Antony & J. Levine, "Reduction with Autonomy", ibid.

[7] Cf. T. Horgan, "Kim on Mental Causation and Causal Exclusion", in J.E. Tomberlin (ed.) op. cit.

However, a second-order functional property is exemplified by the very same sort of thing which possesses its lower-order physical realizer on a given occasion. This is why the problem of causal exclusion does apply and the causal efficacy of the former is threatened by the causal efficacy of the latter.

I now turn to Kim's multiple realization argument revisited. As I said, Kim disagrees with Putnam's celebrated argument that, if a higher-order functional property is multiply realizable by a disjunction of different lower-order physical properties on different occasions, then the former cannot be identical with any of the physical disjuncts and the scientific autonomy of the former is secured. In an earlier paper,[8] Kim argued that if the disjunction of lower-order physical realizers of the higher-order functional property is heterogeneous, then the following dilemma arises: either both the functional property and the heterogeneous disjunction are nomic kinds or neither is. He recommended the latter option on the basis of the following reasoning. Samples of what was once held to be a single mineral kind, jade, turns out to involve two distinct minerals with different molecular structures: jadeite and nephrite. Hence, jade turns out to be a disjunction of two heterogeneous mineral kinds. Hence, "jade is green" does not express a nomic generalization; rather it expresses a conditional with a disjunctive antecedent. Being a disjunction of two heterogeneous mineral kinds, jade is not a projectible property. Not being projectible, it is not a nomic property. In other words, to say that "if something is either jadeite or nephrite, it is green" is to accept the following conjunction: "if something is jadeite, then it's green and if something is nephrite, then it's green". If all our evidence happened to confirm only one of the conjuncts, we should not extend our confidence to the generalization involving the disjunctive antecedent. If so, then psychology cannot be an autonomous science.[9]

---

[8] J. Kim, "Multiple Realization and the Metaphysics of Reduction", in Kim, Supervenience and Mind, op. cit. A version of this argument appears in the book in the section "Multiple Realization Again", pp. 106-12.

[9] I fail to see the force of Fodor's distinction between open and closed disjunctions in response to Kim's argument in his "Special Sciences: Still Autonomous After All These Years", in J.E. Tomberlin (ed.) op.cit.

I do think that in combination with the supervenience argument, Kim's multiple realization argument makes an impressive case against the autonomy of mental properties. The hard question we are left with is: does Kim's reductionism really vindicate mental realism? As I mentioned above, if a mental (e.g., phenomenal) property is not functionalizable, then the choice faced by a physicalist is: epiphenomenalism or eliminativism? But what about functional properties? In one section of chapter 4 (pp. 103-106), Kim entertains the radical view that one ought to trade higher-order functional properties for higher-order functional concepts. This proposal, I think, amounts to giving up reductionism and embracing eliminativism. It strikes me as odd in at least two respects.

First of all, I don't quite I understand the view that there could be functional concepts and no functional properties (p. 104). A second-order property is supposed to be the property of possessing some lower-order physical property meeting some causal conditions. I fail to see what the relevant notion of a second-order concept might be other than the concept of a second-order physical property. However, if there are no second-order physical properties, then what could a second-order concept be? What would it be a concept of? Presumably, the answer is: it is the concept of an uninstantiated property. I wonder whether in retrospect Kim does not himself acknowledge the oddity of his own suggestion when he writes (p. 111): "I also urged that M be viewed as a concept, not a property in the world; however this is to duck the issue — we need a more straightforward answer".

Finally, it would be ironical if by the end of the book Kim ended up agreeing with Davidson's anomalous monist view that we should replace talk of mental properties by talk of mental conceps (or predicates). Of course, unlike Davidson, Kim is not urging that we replace talk of physical properties by talk of physical concepts (or predicates). But giving up mental properties suffices to remove the question: if some physical property of a cause was efficacious in the process whereby it produced its effect, then could its mental property be efficacious too? If so, then the problem of

causal exclusion simply does not arise and Kim's supervenience argument is undermined.