

Jaegwon Kim, *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*
Cambridge, Mass.: MIT Press 1998
ISBN 0-262-11234-5
£19.95 Hardback

As Jaegwon Kim points out in his excellent new book, “reductionism” has become something of a pejorative term in philosophy and related disciplines. But originally (eg, as expressed in Ernest Nagel’s 1961 *The Structure of Science*) reduction was supposed to be a form of explanation, and one may wonder whether it is reasonable to reject in principle the advances in knowledge which such explanations may offer. Nagel’s own view, illustrated famously by the reduction of thermodynamics to statistical mechanics, was that reduction is a relation between theories: theory A is reduced to theory B by formulating “bridge laws” which link the terminology of the theories, and using them to derive A from B. (An additional reductive claim is that A-phenomena are identical with certain B-phenomena—as when the temperature of a gas is identified with its mean molecular kinetic energy—but this kind of identity claim is, strictly speaking, independent of the claim about theories.) Applied to the case of mental states and brain states, a reduction would provide explanatory relations between psychology and neuroscience, normally supplemented with the claim that mental properties are identical with physical properties in the brain.

Such a theory would surely offer an explanation of mind; so why do so many philosophers reject the very idea of reductionism? It is sometimes said that a reductive identity theory denies the existence of mind; but this is a simple mistake. To identify phenomena A and B is to deny neither the existence of A nor of B; on the contrary, the identification presupposes their existence. Despite this, many object that the kind of explanation given by reductionists must inevitably “leave something out”.

Expressed in this way, the objection is fatuous. The mechanical reductive explanation of thermodynamic properties is a genuine explanation: that is, it is an advance in our understanding of the phenomena explained. If there were a parallel explanation of mental phenomena in neuroscientific terms, then this too would be an

advance in our knowledge. The mere fact that, if such an explanation were provided, it would be reductive, should not be a reason for rejecting it.

These confusions aside, one argument which has had a powerful influence in encouraging philosophers of mind to abandon reductionism is the “multiple realization” argument invented by Hilary Putnam. Putnam argued that it is unlikely that all creatures in a common mental state—such as believing that grass is green, or being in pain—will share a common physical state, and if so, there cannot be one physical state or property which is identical with the mental state. Mental properties, in other words, can be “multiply realized” by physical properties, and so are not identical with them.

Many “non-reductive” physicalist philosophers who accept Putnam’s argument have sought to understand the “realization” relation between mental and physical properties in terms of the idea of supervenience. To say that something A supervenes on something else B is to say that any change or difference in A entails a change or difference in B. For example: it is sometimes claimed that moral properties like goodness supervene on non-moral properties. This means that if two actions differ in their moral properties (say, one is good and the other is not) then they must differ in some non-moral way too. Or, in other words, if two actions have all the same non-moral properties, then it cannot be the case that one of them is good and the other one is bad. The idea of supervenience, then, is used to express the sense in which the moral depends on the non-moral. And similar claims are made about the relation between mental properties and the physical properties on which they depend. To say that mental properties supervene on physical properties is to say that two people cannot differ in their mental properties (eg one is in pain and the other is not) without differing in their physical properties. Or, alternatively, if two people have all the same physical properties, then they have all the same mental properties. This claim of mental/physical supervenience has been defended by many non-reductive physicalists.

The idea of supervenience is one which Jaegwon Kim has made his own, and his collection *Supervenience and Mind* (1993) brought together his most important papers on the subject. Sometimes, however, it is hard not to feel that more distinctions are made between kinds of supervenience than there are relevant philosophical problems to solve, and that since the idea ultimately amounts to a kind of (often unexplained) correlation between properties, the mind-body problem will not be illuminated by appeal to supervenience.

It may come as a surprise to readers of *Mind in a Physical World* that Kim now accepts this assessment of the merits of supervenience. In the first chapter we are told that “supervenience is not a metaphysically ‘deep’ relation ... mind-body supervenience *states* the mind-body problem—it is not a solution to it”. Kim has a number of reasons for rejecting his earlier view that supervenience can be a physicalist solution to the mind-body problem: in particular, supervenience is compatible with various forms of dualism, and supervenience itself is in need of explanation.

Kim’s change of mind on this issue may help confirm the suspicions of those who felt that “supervenience studies” was beginning to look like a denegerating research programme. But at the heart of Kim’s new position is scepticism about non-reductive physicalism itself. The reason for this scepticism will be familiar to philosophers of mind: non-reductive physicalism cannot account for mental causation, the fact that mental states have effects in the physical world. The argument for this conclusion is simple but devastating: physicalism is committed to all physical effects having physical causes; non-reductive physicalism denies that mental causes are identical with physical causes; so physicalists who believe in mental causation must accept that some physical effects have distinct mental and physical causes. But these causes cannot each bring about the same physical effect independently of each other, otherwise it would seem a miraculous coincidence that whenever (say) a person’s decision caused their arm to move, events in their brain caused the same thing. So if this kind of “causal overdetermination” is rejected, then either there are

physical effects with no physical causes, or there is no mental causation. But neither of these conclusions is acceptable to a non-reductive physicalist. Kim shows that it does not help to respond to this argument by appealing to the supervenience of the mental on the physical, since a similar version of the argument (which Kim calls the “supervenience argument”) can be constructed which assumes supervenience as a premise.

Kim’s discussion of the various attempts by non-reductive physicalists to explain mental causation has the clarity and simplicity one ought to expect from a philosopher utterly at home in this debate. Vague or purely metaphorical attempts to define relations of “causal relevance” which supposedly differ from causation are adroitly undermined, as are attempts by Tyler Burge and others to deflate the problem by stressing the adequacy of our ordinary mentalistic explanations. As Kim says, that these explanations are adequate is something we knew already; the real question is how their causal claims are consistent with other physicalist assumptions.

The obvious physicalist solution to the mental causation problem is to identify mental and physical causes. If causes are properties, as Kim and others believe, then this involves a return to reductionism, and this is the main message of Kim’s book. But Kim rejects Nagel’s model of reduction, for some of the same reasons that he rejects supervenience as an explanatory theory. Instead, he proposes a functionalist account of the reduction of mental properties, where a mental property is characterized in terms of its typical causes and effects (its causal role) and then identified with the physical occupant of that role. This account is familiar from the 1960s identity theories of D.M. Armstrong—whom Kim credits—and David Lewis—whom, curiously, he does not. The latter omission is especially surprising since Kim’s response to the multiple realization objection is substantially the same as Lewis’s 1969 response.

The possibility of a reduction of mind therefore turns on whether mental properties are functional properties, or as Kim would prefer to put it, whether mental concepts are functional concepts, since he does not believe that “functional/non-

functional” marks a distinction in properties themselves. Although this latter point seems correct, Kim is mistaken when he calls these functional descriptions of properties “relational”, since the typical causes and effects which figure in the functional descriptions of a property need not exist in order for something to have that property. This aside, Kim’s account of the reductionist strategy and its relation to the mental causation problem is one of the best contributions to the recent debate.

Although Kim agrees with many philosophers that intentional mental states can in principle be given functional characterisations, he has doubts about whether conscious mental properties can be “functionalized” in this sense. Here Kim is (by his own admission) less persuasive, and this is one place in his book where Kim appears to lack the courage of his reductionist convictions. (Another is when he concludes his book by giving apparent credit to the non-sequitur that his reductionism leaves the mental with ‘no distinctive role of its own’.) A reduction of consciousness would require, on Kim’s view, a functional conception of consciousness. But it seems to me that the current state of research on the philosophy and science of consciousness does not yet warrant the conclusion that there can be no such conception, tendentious thought-experiments about “zombies” notwithstanding.

Tim Crane
University College London