

The main conclusion of Jaegwon Kim's admirable *Mind and the Physical World* is that the mind-body problem- Descartes problem of explaining how mental causation is possible- has not yet been solved. In particular, non reductive physicalism (NRP), a metaphysical account of the relationship between mental and physical entities that has become increasingly popular among philosophers of mind and that Kim himself once endorsed, is not a viable solution to the problem. I argue here that Kim's arguments against non-reductive physicalism are unpersuasive and suggest that they involve assumptions about causation that are implausible in the light of contemporary physics. When these assumptions are rejected NRP lives.

NRP is a family of views differing by how they understand "reduction" and "physicalism." Following Kim I understand the non-reduction as holding that some events and properties are distinct from any physical events and properties. A necessary condition for physicalism is that mental properties, events, and laws supervene on physical ones. Kim allows various understandings of "supervenience" but I think that physicalism requires at least the claim that any minimal physical duplicate of the actual world is a duplicate *simpliciter*.<sup>1</sup> Some complications aside this means that true mental propositions, e.g. Jaegwon is thinking about sailing, are metaphysically entailed by true physical propositions. Kim says that supervenience is too weak to capture the root idea of physicalism that mental property instantiations depend on physical property instantiations so he adds that the mental *depends* on the physical.<sup>2</sup> One way (but not the only way) in which this dependence might be spelled out is that mental properties are higher order functional properties whose instantiations are realized by instantiations of physical properties. I need to say a little bit about Kim's ontological assumptions. An event is an instantiation of a property by an individual and a time. A mental event is the instantiation of a mental property. Not every predicate expresses a genuine property (disjunctions of properties need not be properties). Kim further suggests that properties are individuated, at least partly, by nomological and causal relations.<sup>3</sup> For physicalism to have content something must be said about the difficult issue of characterizing the physical. Kim's view seems to be that the micro-physical properties of ideal physics

---

<sup>1</sup>Jackson (1998) gives this necessary condition on physicalism.

<sup>2</sup>Two of the reasons that Kim gives for holding that Supervenience is too weak to capture physicalism have to do with features of his formulation of Supervenience. One is that he formulates the doctrine as involving properties. That leaves the possibility that there are laws and other facts that don't supervene on the physical. Another is that he allows that notion of necessity involved in Supervenience is nomological necessity. That allows for the possibility of fundamental bridge laws linking physical and mental properties and the existence of such laws would surely violate physicalism.

<sup>3</sup>Kim seems to endorse Shoemaker's view that properties are individuated in terms of their nomological and causal relations. An apparent consequence of this view is that alleged properties that possess no causal/nomological features are not genuine properties.

are physical.. He also counts as physical properties that are conjunctions and aggregates of micro-physical properties and higher level properties defined over lower-level physical properties. (P115).<sup>4</sup> Since these latter two classes of properties supervene on the micro-properties and laws there is no need to include them in the supervenience base.

Kim's formulation of the problem of mental causation is simple and familiar. Each of the following propositions is plausible:

1. Many (or all) mental events cause physical events (alternatively may mental properties are causally relevant to some physical properties).
2. Mental events are distinct from physical events (the constitutive properties of mental events are not physical properties).
- 3) The physical is nomologically and causally complete and closed.
- 4) Physical events are not pervasively causally overdetermined.

Rejecting 1 would require modifying our view about the mind enormously. It is a last resort. The distinctness of mental properties and events is widely accepted on the basis of the autonomy of psychological explanation and intuitions of multiple realizability and conceivability.<sup>5</sup> That the physical realm is causally/nomologically complete does not mean that physical events *only* have physical events as their causes. If that is how it is understood then the failure of non-reductive physicalism is trivial. Rather it means that if the fundamental laws and the total physical state at  $t$  completely determines the chance (which is 1 or 0 if the laws are deterministic) at  $t$  of an event  $E$ 's occurring at time  $t+k$ . There are no gaps that must be filled in by mental events to account for the occurrence of physical events. Finally, it does seem implausible that every case of mental-physical causation is a case of causal overdetermination, at least if such overdetermination is thought of as like the case in which two assassins each independently kill their victim.

The conflict among 1-4 is obvious. If  $P$  occurs at  $t$  the physical state  $S$  (which may consists of a fusion of many micro-physical events) at  $t-k$  fully accounts for  $P$ . If  $M$  supervenes on  $S$  and is a cause of  $P$  then it appears that  $P$  is overdetermined by  $S$  and  $M$ . If there is no such overdetermination then either  $M$  doesn't cause  $P$  or is identical to some part of  $S$ . If the latter than it is a physical event and so 2 fails. Descartes attempted to resolve it by rejecting the completeness of physics. But this is not an option for physicalists since it involves there being nomological or causal relations that do not

---

<sup>4</sup>Kim thinks that his characterization makes make and biological properties physical. But it doesn't. Being a water molecule is not an aggregate or conjunction of fundamental micro-physical properties but a vast disjunction since water molecules can occupy infinitely many quantum states.

<sup>5</sup> The conceivability intuitions have been deployed in (Chalmers (1996) in attempts to show that certain mental events (properties and facts) don't metaphysically supervene on physical facts and so falsify physicalism. These arguments depend on an inference from conceivability of certain propositions to their possibility. The failure of this inference is established by Balog (2000)

metaphysically supervene on the physical. And, of course, Cartesian dualism has fallen into disfavor in part because it is widely believed on the basis of scientific developments that 3 is true. The non-reductive physicalist's only way out is to deny 4. But Kim thinks that this is especially unattractive for a non-reductive physicalist since it threatens 3. He argues that the best way to understand non-reductive physicalism is one on which the relations between mental events and physical (and other mental) events is not genuine causation at all. Some non-reductive physicalists have responded to the problem by attempting to characterize causation (or causal relevance) in terms of counterfactuals and argue that overdetermination is acceptable. I think that this is the correct approach and will briefly defend it after looking at Kim's arguments against overdetermination..

Kim claims that overdetermination is unacceptable because (p44,45). 1) it is implausible that every physical event that has a mental cause is causally overdetermined, 2) mental causes are dispensable, 3) it may come into conflict with physical causal closure. None of these reasons are persuasive. We need to distinguish the two assassins cases of overdetermination from the kind of overdetermination involved in NRP. An example of such overdetermination is one in which a neurophysiological event N realizes a mental event M (say a desire for a sip of beer) and both cause E (say the subject's hand's moving). In the two assassins case the two causes are metaphysically (and nomologically) independent. In the latter case M depends on N since it is metaphysically (or physically) entailed by it. Cases of the first sort of overdetermination are indeed rare but that provides no reason to think that cases of the second sort are unusual. The reason that we think that the first kind of overdetermination is generally not plausible is that when we posit two *independent* causes C1 and C2 for E we raise the question of why both C1 and C2 and why the events that occurred led to the same effect. But in the case of NRP we don't face these questions since N and M are not independent.<sup>6</sup> Further, mental causes are not really dispensable on the NRP account since the only way M could be dispensed with is by dispensing with P and P may not be dispensable with respect to E. Kim says: "For consider a world in which the physical cause does not occur and which in other respects is as much like our world as possible. The overdetermination approach says that in such a world, the mental cause causes a physical event -namely that the principle of causal closure of the physical no longer holds. I do not think we can accept this consequence: that a minimal counterfactual supposition like that can lead to a major change in the world." (p.45)

This is puzzling. Consider two worlds: in w1 M occurs N doesn't but M is realized by another brain state N\* and both M and N\* are causes of E; in w2 M occurs and causes E but E has no physical cause. Kim seems to be saying that w2 is more similar to the actual world than w1. But why think this? In w1 supervenience and the completeness of physics both hold while in w2 they both fail. In fact it seems obvious that we accept "if M had occurred but had not been realized by N it would have been realized

---

<sup>6</sup>Kim formulates an "exclusion principle" according to which no event has two *independent* causal explanations and which he uses to argue against overdetermination. But it doesn't seem applicable to mental-physical causation in NRP since on it mental and physical events are not metaphysically independent.

by another neural state” and reject the corresponding counterfactual with the consequent “...supervenience and the completeness of physics would have failed.”

It becomes clear why Kim thinks that the reasons he gives against overdetermination should be convincing when he describes what he thinks the right way to describe the situation if NRP holds.

“However, if we understand the difference between genuine, productive and generative causal processes, on the one hand, and noncausal regularities that are observed because they are parasitic on *real (sic)* causal processes, we are in a position to understand the picture recommended by (viii) (i.e. by NRP). In the case of the supposed M-M\* causation the situation is rather like a series of shadows cast by a moving car. The moving car represents a genuine causal process, but the series of shadows it casts, however regular and law-like it may be does not constitute a causal process” (p45)

Kim is thinking of causation as a relation in which the cause generates or produces the effect. If both N and M each produce E then they do so independently. So if N produces E then M is dispensable as a producer of E. If M produces E then if we just take N away M would still produce E thereby violating the causal completeness of the physical. If we reject overdetermination then we will have to say that M doesn't produce M\* (or anything else) but is like the shadow of the car. I will return to the issue of whether this analogy is apt and whether we should understand causation as production at the conclusion of these comments. So Kim's argument against overdetermination, as he makes clear, depends on his conceiving of causation on the model of one event producing another.

Under the heading “free lunches” Kim discusses counterfactual accounts of causation as a way for NRP to have one's causal cake and eat it too. The basic idea is that C causes E if C and E are occurring distinct events and  $-C \succ -E$ .<sup>7</sup> The counterfactual account countenances overdetermination. It is clear that there are situations in which we would be willing to assert both  $-N \succ -E$  and  $-M \succ -E$ . On the most developed account of counterfactuals, David Lewis' (1986), counterfactual overdetermination is compatible with distinctness and completeness. So why isn't this a way in which NRP can both have and eat its cake? Kim gives two reasons. One is that the counterfactual account of causation is inadequate as an account of causation. The other is that Lewis account of counterfactuals-which may solve some of problems he raises for the counterfactual account- involves “heavy metaphysical assumptions.” (P 130 fn19).

Kim's main objection is that the counterfactual account doesn't distinguish epiphenomenal relations from genuine causation. Let S1 be the location of the car's shadow at t1 and S2 at t2 and C1 and C2 the corresponding locations of the car. The counterfactuals  $-S1 \succ -S2$  and  $-S1 \succ -C2$  may strike us as

---

<sup>7</sup>Lewis' proposes a condition that is necessary as well as sufficient. C causes E iff there is a chain of counterfactual dependencies from C to E. There are well known problems with the necessity of the account due to pre-emptive causes. Lewis has recently modified his account greatly to deal with them. But whether or not this is successful if the condition is sufficient and it could be shown that it permits mental causation that is compatible with NRP Kim's objections would be answered.

true though there is no causal relation between S1 and either S2 or C2. Further, it seems to Kim that the supervenience of M on N is compatible with epiphenomenalism with respect to E (indeed Kim thinks this is the case if M is distinct from N) and yet in this situation  $\neg M \triangleright \neg E$ . So the counterfactual account can't distinguish epiphenomenalism from mental causation. Kim also observe that  $\neg E \triangleright \neg E^*$  may be true when E doesn't cause E\* but both are effects of a common cause and that  $\neg E \triangleright \neg C$  may be true when C causes E. If these objections are correct than the counterfactual account fails.

I think these objections can be overcome by combining a Lewis' account of counterfactuals with a slightly more complicated account of causation. Roughly, on Lewis' account  $A \triangleright B$  is true iff at all the most similar to the actual world at which A holds B also holds. Lewis assumes that worlds can be characterized in terms of their complete histories of instantiations of fundamental categorical facts and their laws. Counterfactual and causal relations supervene on these.<sup>8</sup> World similarity is evaluated in terms of similarity in history and similarity in conformity to laws of the actual world. I am not sure what "heavy" metaphysical assumptions Kim has in mind. It is widely believed that Lewis' account of counterfactuals depends on his version of possible world realism. But this is a mistake. For Lewis the truth makers of counterfactual statements are the actual facts and laws. Possible worlds are merely a device for spelling out their semantics and there is no commitment to any particular view of worlds. Now it is important to keep in mind that in evaluating Lewis' counterfactual account of causation one must employ counterfactuals characterized by Lewis and not merely appeal to intuitions about counterfactuals. Lewis argues that his counterfactuals (and so the account of causation in terms of these counterfactuals) can handle the problem of the direction of causation and the problem of effects of a common cause.<sup>9</sup> In order to handle the problem concerning epiphenomenalism for higher level events we need to complicate Lewis' account. Say that M causes E if  $\neg M \triangleright \neg E$  and there is no lower level event that preempts this relation that is not itself preempted by M.<sup>10</sup> N preempts M with respect to E if  $\neg N \triangleright \neg E$  and  $N \& \neg M \triangleright E$ . The intuitive idea is that M causes E if E counterfactually depends on N and but doesn't counterfactually depend on M as long as N occurs. This condition clearly makes the shadow epiphenomenal wrt the car's position and the future shadow. Both the mental epiphenomenalist and the

---

<sup>8</sup> Lewis assumes that fundamental properties are categorical (laws and causal relations are not essential to fundamental properties so his account of counterfactuals and causation is not circular. He further assumes that laws supervene on the categorical property instantiations. While I like this Humean view it is not part of my defense of the counterfactual account of causation.

<sup>9</sup>In fact Lewis is wrong about this for reasons having to do with statistical mechanics. (Elga ms.; Albert and Loewer ms.). However, this is not relevant to my defense of the role of counterfactuals in mental causation since there are repairs that can be made in the Lewis account to avoid the problems.

<sup>10</sup>We also need to include that the physical states that are the supervenience bases for M and E are connected by fundamental dynamical law. A proposal along these lines is found in LePore and Loewer (1987), (1990)

interactionist dualist hold that mental events are (at most) connected by bridge laws to neurophysiological states.<sup>11</sup> It follows that N&-M is metaphysically possible. Suppose that -M>-E so M is a putative cause of E. The epiphenomenalist holds that it isn't a genuine cause since N preempts M while the interactionist dualist holds that M is not preempted and is a genuine cause. The non-reductive physicalist holds that there are mental events M that are putative causes that are not preempted by events that they themselves don't preempt. The existence of such mental events is compatible with non-reductive physicalism.

I don't expect Kim to be persuaded by my counterfactual defense of causation because he considers it, at best, causation lite as compared to causation as production. But causation as production fits ill with contemporary physics. Russell (1912) famously said that causation so understood "is a relic of a bygone age, surviving, like the monarchy, only because it is erroneously supposed to do no harm." What Russell had in mind is that the fundamental laws and facts of physics do not mention causality. Further, those laws relate the totality of the physical state at one instant to the totality at later instants.<sup>12</sup> They do not single out parts of states at different times as being causally related. If one wants to think of causation as production then one has to say that the entire micro physical state at t produces the state at t'. In discriminating among the events in the states prior to the occurrence of E we are interested in the extent to which, given the laws, these events *make a difference* to E's occurrence. This is the idea that counterfactual accounts attempt to capture. It may be that Lewis' particular account (or my development of it) doesn't succeed in explicating the making a difference concept of causation, but we have every reason to believe that it can be explicated in terms of the fundamental laws and non-causal facts since we have every reason to think that they make up the world. On the other hand, the production conception of causation seems to involve facts that fail to supervene on the fundamental laws and facts. It seems to allow the existence of two universes that match perfectly in fundamental laws (of the sort posited by contemporary physics) and facts and yet differ in which events "produce" which others.<sup>13</sup> This indeed is a heavy metaphysical assumption.

Propositions 1-4 are inconsistent. NRP holds the first three and rejects 4. The kind of overdetermination it requires is innocuous when causation is understood in terms of counterfactuals. Kim thinks of causation as production and so finds 4 more plausible than the conjunction of 1-3. But giving up one of those, as Kim makes clear, leads to dualism or epiphenomenalism or reductive physicalism. This seems to me yet another reason to reject the production conception of causation. A free lunch is preferable to indigestible metaphysics.

---

<sup>11</sup>Such bridge laws do not supervene on the physical laws and state.

<sup>12</sup>Or a little more correctly a region of the state at time t is related by law to the states in its backward light. This point which is often neglected by philosophers is forcefully made in Latham (1987) and Field (ms.)

<sup>13</sup>Tooley (19xx) who holds a production conception of causation argues for this very possibility.

Barry Loewer  
Rutgers

References:

- Balog, Katalin (2000) "Conceivability, Possibility, and the Mind-Body Problem", *The Philosophical Review*.
- Chalmers, David (1996) *The Conscious Mind*
- Field, Hartry (ms.) "Causation in a Physical World"
- Jackson, Frank (1998) *From Metaphysics to Ethics* Oxford: Clarendon Press.
- Latham, Noa (1987) "Singular Causal Statements and Strict Deterministic Laws", *Pacific Philosophical Quarterly* 68:29-43
- Lepore, Ernest and Loewer, Barry (1987) "Mind Matters" *The Journal of Philosophy* (1987) 630-42.
- Lepore, Ernest and Loewer, Barry (1991) "More on Making Mind Matter More" *Philosophical Topics*
- Lewis, David (1986) *Philosophical Papers II*. Oxford: Oxford University Press
- Russell, Bertrand (1912) "On the Notion of Cause". *Proceedings of the Aristotelian Society* 13: 1-26.
- Tooley, Michael *Causation*