

Mental Causation Through Constitution

Diana Mertz Hsieh (diana@dianahsieh.com)

Mental Causation and Intentional Action (Phil 6300, Hanna)

6 March 2003

In philosophy of mind, many of the debates in recent decades over the nature of mental causation have been framed in terms of two related problems formulated by Jaegwon Kim: the Explanatory Causal Exclusion Problem and the Supervenience Causal Exclusion Problem. Both cast doubt on the possibility of a coherent nonreductive materialist account of mental causation. The Explanatory Causal Exclusion Problem indicates that the nonreductionist appeal to two “complete and independent” causal explanations of minded behavior, one mental and one physical, contradicts the materialist commitment to the Principle of the Causal Closure of the Physical (Kim 1989, 79). The Supervenience Causal Exclusion Problem suggests that the nonreductive materialist advocacy of supervenience renders downward (mental to physical) causation superfluous at best and incoherent at worst. Due to the strength and clarity of Kim’s arguments, any theory of mental causation must directly grapple with these two problems.

Derk Pereboom’s essay “Robust Nonreductive Materialism” attempts to do precisely that through his formulation of a “robust” nonreductive materialism. He tackles both the Explanatory and Supervenience Causal Exclusion Problems by asserting a causal relationship of constitution between the mental and the physical, justified by an argument for token multiple realization. The basic question of this paper is thus whether Pereboom’s solutions are adequate to the problems identified by Kim.

Standard Nonreductive Materialism

Nonreductive materialism, “the received view” of contemporary philosophy of mind, is largely an attempt to accommodate identity theory to Putnam’s multiple realization argument (Kim 1993, 339, 341). As such, three basic theses are integral to the position: physical monism, nonidentity of the mental and physical, and strong supervenience.¹

Thesis 1: Physical monism. According to physical monism, everything real is physical (Kim 1993, 266). Applied to causality, physical monism requires the principle of the causal closure of the physical, which states that all physical effects must have (sufficient) physical causes (Kim 1993, 280). This principle serves to exclude the wholly and exclusively mental causes of physical effects found in Cartesian substance dualism. However, due to the nonidentity of the mental and physical (Thesis 2), nonreductive materialists combine their physical monism with a weak property dualism in which mental properties may exist dependently upon physical properties (Kim 1993, 267). These mental properties are supposed to do “real causal work” within nonreductive materialism (Kim 1993, 279).

Thesis 2: Nonidentity of the mental and physical. Nonreductive materialists explicitly reject the reductive claim that the mental is identical to the physical on the basis of the multiple realization argument. (After all, the mental and physical cannot be strictly identical if the same mental properties may be realized in a wide variety of physical systems.) Instead of identity, nonreductive materialists advocate the strong supervenience of the mental upon the physical

¹ The central principles of nonreductive materialism can be (and are) formulated in a variety of ways, given that the position consists of a constellation of interconnected ideas. Thesis 2, in particular, could probably be replaced with non-identity, property dualism, multiple realizability, and so on.

(Thesis 3). The rejection of identity through multiple realization has two basic consequences for nonreductive materialism. First, many nonreductive materialists are also functionalists, such that mental states are identified with abstract relational states rather than particular intrinsic physical states. Second, nonidentity is regarded as justifying the mental as an “irreducible, autonomous domain” from the physical (Kim 1993, 267). This autonomy of the mental is the basis for the nonreductive materialist commitment to “compatibilist” explanations, under which minded behavior may be explained as the product of either physical or mental causes, each of which is sufficient to explain the behavior, but neither of which is reducible to the other.

Thesis 3: Strong supervenience. In supervenience, any mental difference or change requires an underlying physical difference or change. Strong supervenience additionally asserts a relationship of necessary dependence between the mental and the physical, such that “what mental properties a given thing has depends on, and is determined by, what physical properties it has” (Kim 1998, 11). Strong supervenience is supposed to be materialism “free of reductionist commitments,” but only in the limited sense that the relationship is not one of identity (Kim 1993, 275).

The difficulty of nonreductive materialism lies in explaining mental causation—particularly mental to physical causation—within the framework of these three theses. If Kim’s formulations of the Explanatory and Supervenience Causal Exclusion Problems are correct, then nonreductionism is simply incompatible with materialism (Kim 1993, 267).

The Explanatory Causal Exclusion Problem

One appealing consequence of nonreductionism is its preservation of our common-sense explanations of human and animal behavior in terms of “folk psychology.” Because the mental is not identical to the physical, our perceptions of sunsets, thoughts about tax cuts, desires for bologna sandwiches, and other mental states cannot be reduced to “nothing but” purely neurological or microphysical states. At the same time, the causal closure of the physical domain under materialism requires all physical actions (including minded behavior) to have a sufficient physical cause. The combination of these two lines of thought in nonreductive materialism generates a commitment to sufficient parallel mental and physical causal explanations of the same behavior. So grasping a set of keys might be fully explained as the effect of physical patterns of neural activation—or fully explained as the effect of a mental desire to drive to the store to buy a gallon of milk. These explanations are supposed to be “simultaneous nomic equivalents” such that “one constitutes an explanation... if and only if the other does” (Kim 1989, 83). In short, the compatibilist explanations of nonreductive materialism seem to combine a respect for the integrity of the mental with a devotion to the fundamentality of the physical.

Unfortunately for the nonreductive materialist, Jaegwon Kim’s Explanatory Causal Exclusion Problem casts significant doubt upon even the possibility of such compatibilist explanations.² His examination of the five possible relationships between causal explanations, in fact, lends strong support to his principle of causal exclusion that “no event can be given more than one *complete* and *independent* explanation” (Kim 1989, 79).

Kim’s argument for the principle of causal exclusion starts with a seemingly simple and innocuous question: What are the possible relationships between two causal explanations of the same event? (Kim 1989, 85). More particularly, Kim asks: Is the nonreductive appeal to “two

² Kim’s entire formulation of the Explanatory Causal Exclusion Problem, we should note, explicitly depends upon his commitment to causal and explanatory realism (Kim 1989, 74).

distinct events [as] nomologically equivalent causes or explanations of the same phenomena” a genuine option for a materialist? (Kim 1989, 86). Based upon his response to the first question, Kim answers the second question with a clear “no.” He argues that compatibilist explanations introduce an “instability” into our beliefs that “generates a strong pressure to find an acceptable account of the relationship” between the two explanations (Kim 1989, 85). Equilibrium is achieved only when the “two explanations (or causes) in effect collapse into one” or when we discover one explanation to be “dependent on, or derivative from, the other” (Kim 1989, 86). Kim supports this characterization with an enumeration of the five possible relationships between two causal explanations of the same event: identity, supervenience, jointly sufficiency, interaction, and overdetermination. None of these relationships, he argues, yields “complete and independent” explanations—at least not in a way palatable to nonreductive materialism (Kim 1989, 79).

So what are the possible relationships between two explanations (E1 and E2) citing two apparently distinct, complete, and independent causes (C1 and C2) of the same event? As an example throughout these five cases, let’s consider the causal explanations of the placement of six apples in a bowl.

In an *identity* relationship between E1 and E2, C1 and C2 are actually one and the same cause (Kim 1989, 89). Thus E1 might be “Fred placed the six apples in the bowl” and E2 might be “Mr. Rogers placed the six apples in the bowl.” These explanations seem mutually exclusive until we realize that Fred is Mr. Rogers.

In a *supervenience* relationship, C1 and C2 are “distinct,” but C1 is nevertheless “dependent” upon C2 (Kim 1989, 90). For example, E1 might be “Dominique dropped six apples into the bowl” and E2 might be “A biological gripping device holding an apple was released four inches above the bowl six times.” These two explanations are not equivalent, as they convey different information. Nevertheless, E1 can be reduced to the “more fundamental causal process” of E2 (Kim 1989, 90).

Under *joint sufficiency*, C1 and C2 are not themselves sufficient causes, but “each is an indispensable component of a sufficient cause” (Kim 1989, 90). Thus E1 might be “Hagrid put two apples in the bowl” and E2 might be “Harry put four apples in the bowl.” Neither of these explanations alone fully explains the presence of the six apples, but together they do.

With *interaction*, C1 and C2 are “different links in the same causal chain” leading to E (Kim 1989, 90). So E1 might be “Mary Ann put 12 apples in the bowl” and E2 might be “Gilligan removed six apples from the bowl.” When put into the proper sequence, these two incomplete explanations form one complete explanation.

In a case of *overdetermination*, C1 and C2 are both sufficient causes each “belonging to two distinct and independent causal chains” (Kim 1989, 91). So E1 might be “Romeo placed apples from his bag of ten apples into the bowl until it held exactly six” and E2 might be “Juliet placed apples from her bag of ten apples into the bowl until it held exactly six.” If Romeo and Juliet were placing apples into the bowl at the same time, the event would be overdetermined, for if either person had been absent, the bowl would still contain six apples.

At first glance, the only relationship of these five seemingly capable of accommodating the parallel mental and physical explanations of nonreductive materialism is overdetermination (Kim 1989, 92). Identical explanations aren’t really different explanations at all. A supervenient explanation is fundamentally dependent upon the basal explanation. Jointly sufficient and interactive explanations must be combined to form a single complete explanation. In contrast to these four, overdetermined explanations may at least be justly thought of as “exempt” from the

principle of explanatory exclusion (Kim 1989, 92).³ This understanding of the relationship between compatibilist explanations of minded behavior is supported by the fact that, like in overdetermination, each explanation is regarded as sufficient (Kim 1989, 86). However, in facing this unpalatable option, the nonreductive materialist may emphasize that the compatibilist explanations are not truly independent causal chains as in classic cases of overdetermination, but rather “nomic equivalents” which “co-occur as a matter of law” (Kim 1989, 86). But that relationship, Kim argues, is virtually indistinguishable from the equally unpalatable option of joint sufficiency (Kim 1989, 86).

In the end, neither overdetermination nor joint sufficiency constitutes a genuine option for materialism, as both violate the central materialist commitment to the principle of the causal closure of the physical. Overdetermination violates causal closure because “in the counterfactual situation in which the physical cause had not occurred, the mental cause would have occurred and caused the physical effect” (Kim 1993, 281). Joint sufficiency of the mental and the physical violates closure because “it regards the mental event as a necessary constituent of the full cause of a physical event” (Kim 1993, 280). This conflict between compatibilist explanations and causal closure is the central difficulty of the Explanatory Causal Exclusion Problem for the nonreductive materialist.

The Twin Requirements Problem of Supervenience

As concerns mental causation, only relationships of identity and supervenience between mental and physical causal explanations are consistent with the causal closure of the physical domain. As what makes reductive materialism reductive, a relationship of identity is clearly not a viable option for nonreductive materialism. And in keeping with the strong supervenience of Thesis 3, the hopes of nonreductive materialism are clearly vested in some form of supervenience through which mental properties are “in some substantive sense dependent on, or determined by” physical properties but “not reducible to them” (Kim 1993, 275).

But yet again, Kim offers strong reasons to doubt that any form of supervenience can fulfill the needs of nonreductive materialism, namely “twin requirements” of nonreduction and dependence (Kim 1993, 276). Kim argues that the “main difficulty” is that

if a relation is weak enough to be nonreductive, it tends to be too weak to serve as a dependence relation; conversely, when a relation is strong enough to give us dependence, it tends to be too strong—strong enough to imply reducibility (Kim 1993, 276).

Kim makes his case for this conflict through an examination of the results of opting for global versus local supervenience. Global supervenience, he admits, “does not entail the existence of psychophysical laws” and thus is “consistent with the nomological irreducibility of the mental to the physical” (Kim 1993, 277). However, such global supervenience offers only the weakest of dependence relationships between the mental and the physical, as it allows that a change in a single atom in the depths of space could radically alter the mental features of the world, such as by redistributing phenomenal consciousness from humans to rocks (Kim 1993, 277). Local reduction, on the other hand, offers the necessary relationship of strong dependence between the mental and the physical—but sacrifices irreducibility in the process (Kim 1993, 278). As Kim notes, talk of “specific correlations and dependencies between specific psychological and physical properties” is really talk of psychophysical laws with which “there is always the threat,

³ However, as Kim notes, each explanation in a case of overdetermination may also be thought of as “incomplete” in isolation (Kim 1989, 92). In some significant sense, whether an isolated overdetermined explanation is considered complete or not depends upon the abstractness of the description of the event.

or promise, of psychophysical reduction” (Kim 1993, 278-9). As such, the nonreductive materialist may be able to opt for either dependence or irreducibility in his supervenience relationship—but not both.

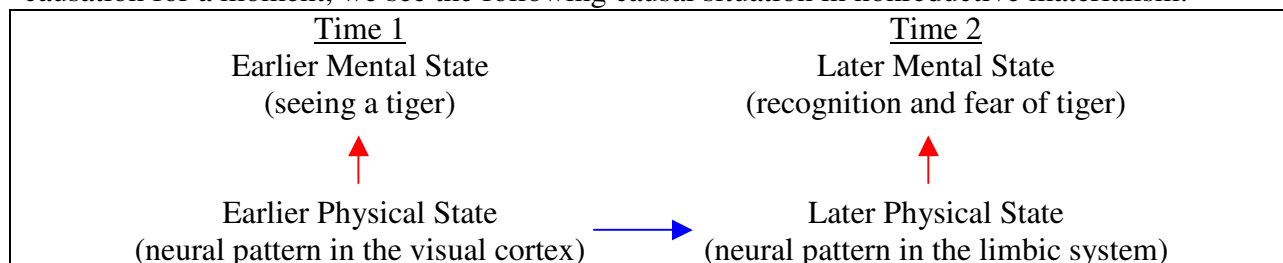
However, the conflict between these two dimensions of supervenience is not the only problem faced by the nonreductive materialist opting for a supervenience relationship between the mental and the physical. Even if a perfectly balanced supervenience were to be elucidated, the nonreductive materialist would still face Kim’s Supervenience Causal Exclusion Problem.

The Supervenience Causal Exclusion Problem

The Supervenience Causal Exclusion Problem arises out of a basic conflict between the combination of “upward determination” and “downward causation” in nonreductive materialism (Kim 1993, 353). In essence, the combination of strong supervenience with causal closure seems to yield to an epiphenomenalist account of mental causation, one that excludes the much-desired rightward (mental-to-mental) and downward (mental-to-physical) causation. Additionally, any insistence upon rightward and downward mental causation within the framework of strong supervenience pushes nonreductive materialism back into either overdetermination or joint sufficiency—thus raising the specter of the Explanatory Causal Exclusion Problem again.

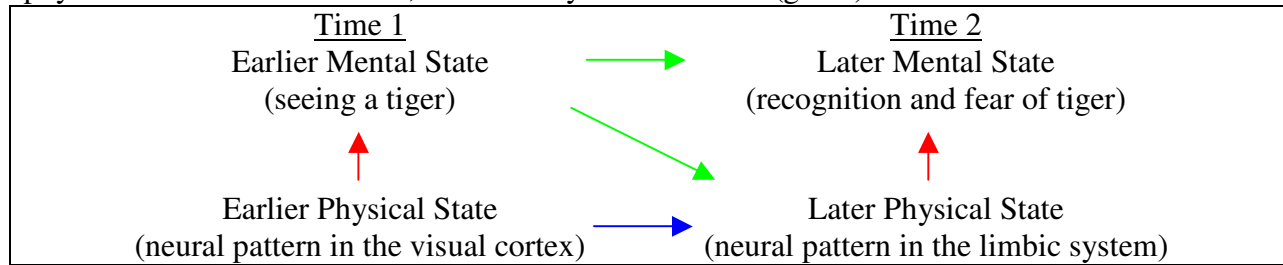
In nonreductive materialism, the mental strongly supervenes on the physical, such that “mental properties are in some robust sense *dependent on* or *determined by* physical-biological properties” (Kim 1993, 340). Consequently, “what objects have which *physical properties* (including relations) determines what *mental properties* these objects have” (Kim 1993, 341). At the same time, nonreductive materialism is committed to all of the three basic forms of mental causation: physical-to-mental, mental-to-mental, and mental-to-physical. However, due to nonreductionism, such mental causation cannot be reduced to nothing but underlying physical-to-physical causation. Rather, “mental properties bring with them *new causal powers, powers that no underlying physical-biological properties can deliver*” (Kim 1993, 350). As such, hunger must somehow causally contribute to eating, sadness must somehow be causally responsible for crying, joy must in some way cause jumping, and so on. As Kim has noted, these views bring nonreductive materialism very close to emergentism (Kim 1993, 344).

Although the combination of strong supervenience and full mental causation is appealing, we might wonder precisely how well these two ideas integrate. In particular, if the nature of the mental is wholly determined by the nature of the physical, how could the mental possibly cause changes in the physical? The difficulty becomes more evident when sketched in visual form over time (Kim 1998, 149-51). Leaving aside mental-to-mental and mental-to-physical causation for a moment, we see the following causal situation in nonreductive materialism:



The (red) upward arrows from the physical states to the mental states represent the supervenience relationship of upwards determination. The (blue) rightward arrow represents the earlier physical state as a sufficient cause of the later physical state (in accordance with the causal closure of the physical). For the nonreductive materialist, this picture is incomplete and

unsatisfactory, as the mental is doing no causal work; it is merely an epiphenomenon of physical causal processes. The earlier mental state ought to be causally influencing the later mental and physical state in some fashion, as shown by the additional (green) arrows in this chart:



But, Kim asks, why bother adding such mental-to-mental and mental-to-physical causation when we already have fully sufficient causal explanations of all the events without them? (Kim 1993, 354). In other words, Kim notes:

Given that [the earlier physical state] is a sufficient cause of [the later physical state], how could [the earlier mental cause] *also* be a cause, a sufficient one at that, of [the later physical state]? What causal work is left over for [the earlier mental state], or any other mental property, to do? (Kim 1993, 354).

This is, in a nutshell, the big question of the Supervenience Causal Exclusion Problem.

More interestingly, insisting upon adding mental-to-mental and mental-to-physical causation as in the second figure seems to trap the nonreductive materialist back into the Explanatory Causal Exclusion Problem, in that the causal configuration then begins to look like either overdetermination or joint sufficiency again. As such, Kim has laid a clever trap for the nonreductive materialist with his Explanatory and Supervenience Causal Exclusion Problems, such the solution to one problem seems to entail the emergence of the other.

Pereboom’s Robust Nonreductive Materialism

In his article “Robust Nonreductive Materialism” Derk Pereboom explicitly attempts to formulate a version of nonreductive materialism “robust” enough to avoid both the Explanatory and Supervenience Causal Exclusion Problems (Pereboom 2002, 499). His basic strategy is to extend multiple realization to higher-level tokens in order to justify the idea of the mental as constituted by the physical, but not identical to it. The resulting approach to mental causation seems to avoid the worst of the two Exclusion Problems, but only by cozying up to identity and weakening any substantial commitment to nonreduction.

The standard version of multiple realization found in nonreductive materialism is concerned with the multiple realizability of mental state *types* (such as pains, itches, and tickles) in a variety of physical bases. Pereboom’s basic contribution is his extension of multiple realization to mental state *tokens* and to higher-level tokens generally, on the grounds that “higher-level token entities are ... not identical to their realization bases” because they can be multiply realized much in the same way as types (Pereboom 2002, 503). Pereboom argues this point largely through his example of the ship of Theseus, in which the replacement of a few planks does not alter its status as the same higher-level token entity. He writes,

The ship of Theseus is not identical with its current token microphysical realization base, for it would have been the same token ship had the token microphysical realization been slightly different, and it will be the same ship when this microphysical realization if fact changes—the ship is in this sense token multiply realizable (Pereboom 2002, 503).

Additionally, because the ship of Theseus itself is token multiply realizable, so are its token causal powers (Pereboom 2002, 503). For Pereboom, this extension of multiple realization to higher-level tokens straightforwardly applies to mental properties and mental causal powers.

Thus armed with higher-level token multiple realization, Pereboom develops a superstrong form of supervenience of the mental upon the physical, namely that of constitution. Under constitution, mental properties, events, and causal powers are regarded as “wholly constituted” by microphysical properties, events, and causal powers (Pereboom 2002, 500, 504). Such constitution is so close to identity that it borders on reductionism; according to Pereboom “the token causal powers [of the mental] would be ‘nothing over and above’ the token causal powers of [the physical]” in a relationship of constitution (Pereboom 2003, 504). Nevertheless, the strict reductionism of strict identity is avoided through multiple realization, as well as through an appeal to intrinsic mental properties (Pereboom 2002, 516).

Within the framework of the constitution relationship, Pereboom attempts to avoid the Supervenience Causal Exclusion Problem largely by distancing his robust nonreductive materialism from emergentism. His admission that mental causal powers are in some significant sense “absorbed” or “swallowed up” by the causal powers of the physical base in constitution runs counter to emergentism (Pereboom 2002, 504). He explicitly argues that constituted mental causal powers do not have the emergentist “power to produce deviations from the ordinary microphysical laws” (Pereboom 2002, 510-11). And he stresses that robust nonreductive materialism denies the central emergentist principle that some higher-level properties are not predictable from their basal conditions. Rather, the relationship of constitution allows for as much “theoretical predictability” of higher-level properties as found in identity (Pereboom 2002, 509). To the extent to which the Supervenience Causal Exclusion Problem depends on nonreductive materialism’s proximity to emergentism, Pereboom has made his case well. But perhaps more significantly, Pereboom denies Kim’s implicit presumption that less-than-completely fundamental explanations are dispensable. He writes,

Suppose that M instantiates an irreducibly mental causal power, and that the activation of this irreducibly mental causal power brings about P*. Accordingly, M will cause P*, and M’s causing P* will not be not identical with P’s causing P*, and thus there is a significant sense in which M is indispensable as a cause of P*. Given a conception according to which explanations track and express the causal relations in which causal powers participate, the explanation of P* in terms of M will be *bona fide* and indispensable. If one were to claim instead that genuine explanations account for phenomena only in terms of the most fundamental conditions that cause them, then explanations in terms of mental causal powers might well be dispensable, while only microphysical explanations would survive. But this is not a notion of explanation that we are constrained to accept (Pereboom 202, 511).

On the whole, Pereboom seems to be denying the basic picture drawn by Kim’s Supervenience Causal Exclusion Problem, arguing instead for a much tighter connection between the mental and the physical.

As for the Explanatory Causal Exclusion Problem, Pereboom sees his robust nonreductive materialism as avoiding “competition between [mental and physical] explanations” due to a “constitutional coincidence” in which “two sets of causal powers...are constituted from precisely the same stuff” (Pereboom 2002, 505). Such constitutional coincidence does stop just short of identity because of higher-level token multiple realization. Nevertheless, it allows for the possibility of “two causal explanations for one event that do not exclude each other and at the same time do not reduce to a single explanation” (Pereboom 2002, 505).

In some formal sense, Pereboom does seem to have provided reasonable solutions to both the Explanatory and Supervenience Causal Exclusion Problems. Nevertheless, there is a substantial sense in which he has given up nonreductionism in the process. In particular, Pereboom carves the relationship of constitution out of the logical space that only identity used to occupy. Consequently, identity is very strictly limited to perfect sameness at the microphysical level, to the point that the ship of Theseus is not identical to itself from one

moment to the next if a single atom is added or subtracted. Thus Pereboom even admits in his initial discussion of token multiple realizability that

there is some notion of sameness by which the ship is the same thing as its current microphysical realization—a notion that abstracts from any temporally extrinsic or modal properties. But it is sufficient for absolute non-identity that A and B differ in their temporally extrinsic or modal properties (Pereboom 2002, 503).

If such absolute identity is the only form of identity and only foundation of reductionism possible in philosophy of mind, then reductive materialism seems quite impossible. Nevertheless, Pereboom's robust nonreductive materialism is nonreductive at least in name, as it endorses multiple realizability and supervenience. But, we must wonder, is it at all nonreductive in spirit?

Concluding Thoughts

Throughout his discussions of the Explanatory and Supervenience Causal Exclusion Problems, Kim argues that an identity relationship between the mental and the physical is the only viable form of materialism. Although Pereboom's robust nonreductive materialism attempts to prove Kim wrong on that score, the close proximity of constitution to identity seems to leave Kim with the upper hand in the debate.

References

- Kim, Jaegwon. 1989. "Mechanism, Purpose, and Explanatory Exclusion." In *Philosophical Perspectives*, 3.
- Kim, Jaegwon. 1993. *Supervenience and Mind*. Cambridge: Cambridge UP.
- Kim, Jaegwon. 1998. *Philosophy of Mind*. Boulder: Westview Press.
- Pereboom, Derk. 2002. "Robust Nonreductive Materialism." In *The Journal of Philosophy*. October, Vol XCIX, No 10.