

MENTAL CAUSATION AND THE CAUSAL COMPLETENESS OF PHYSICS

WILSON MENDONÇA
Universidade Federal do Rio de Janeiro

Abstract

The paper takes issue with a widely accepted view of mental causation. This is the view that mental causation is either reducible to physical causation or ultimately untenable, because incompatible with the causal completeness of physics. The paper examines, first, why recent attempts to save the phenomena of mental causation by way of the notion of supervenient causation fail. The result of this examination is the claim that any attempted specification of the most basic causal factors which supposedly underlie a causal transaction cannot account for the counterfactually necessary connections with the effect in question. By contrast, the specification of these factors at a higher-level would allow establishing such connections. The paper closes with a discussion of how this view of autonomous higher-level causation grounded on counterfactual relations can be made compatible with the physicalistic commitment to a complete specification of the particular causes of any physical effect exclusively in physical terms.

1. Supervenience

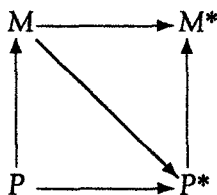
Nonreductive materialists in philosophy of mind have been looking for a way to reconcile physicalistic commitments with a view of mental properties as dependent on, but also essentially distinct from physical properties. The idea that a supervenience relation might hold between physics (conceived of as the most basic science) and the mentalistic discourse (taken as a "special science") seemed to fill the bill. It was once a widely shared conviction that a suitably defined notion of supervenience would make it possible for philosophers of

mind to preserve their materialism while holding on to the “autonomy of psychology” as an irreducible account of the causal relations connecting mental properties. This consensus has been shattered in the meantime by a number of powerful arguments—one of the most prominent being the overdetermination argument (Kim 1989, 1993a and 1993b)

The overdetermination argument challenges the very possibility of mental causation, i.e. the objective relation that supervenience was supposed to secure. It seeks to show that given the causal efficacy of physical properties, which everyone takes for granted, there is no causal role left for supervenient properties to play. The suggestion is that, if mental properties merely supervene on physical properties, it is unlikely that we will be able to find a proper place for them in the causally structured world. If the argument is cogent and valid, no account of mental-to-physical causation is possible that does not flout the materialistic assumption of the causal completeness of physics. The conclusion is that physicalistic commitments to mere dependence or supervenience relations have to be converted into commitments to type-identities between mental properties and physical properties.

I will take issue with this view. My first main thesis is that the overdetermination argument is not as conclusive as it seems. The crucial difficulties concerning mental causation, especially mental-to-physical causation, so I will argue further, find solution in a metaphysical framework that draws on the notion of supervenience and accepts the causal completeness of physics, upon providing an independently justified interpretation of the latter.

2. Overdetermination



Consider the case, where an instantiation of a mental, supervenient property M causes the instantiation of another mental property M^* . An illustration of this would be a mental phenomenon causing another mental phenomenon. As materialists, supervenientists (as we may call them) must assume that the appearance of supervenient properties depends on the presence of appropriate basal conditions. So we have for the mental property M^* a determining physical property P^* . The counterfactual implication of M 's claim to being a cause of M^* says that M^* could not have been instantiated, if M had not been instantiated on this occasion. The determination relation between P^* and M^* , on the other hand, implies that unless P^* were present on this occasion, M^* could not have been instantiated. These two conditions cannot be independent from another. A plausibly coherent description of the situation seems to be: the instantiation of M causes the instantiation of M^* by causing the instantiation of P^* in the first place, the later instantiation determines then the instantiation of M^* .

The first part of this description entails, of course, mental-to-physical causation, a relation objectively connecting a higher-level phenomenon (as the cause) to a lower-level phenomenon (as the effect). To this the supervenientist is committed. Under this line of thought, the causal role of M in the process by which P^* is brought about cannot be entirely "preempted" by any physical property. However, we do have good reasons to assume that the physical preempts the mental. These reasons are derived from the assumption of the causal completeness of physics: the instantiation of P^* has as its cause a physical phenomenon. Hence the purported distinctness of supervenient causal powers results in the uncomfortable supposition that physical phenomena underlying mental phenomena are systematically overdetermined. The instantiation of P^* has two distinct causes, a physical cause and a mental one. What causes discomfort is the fact that the joint operation of two causes, each one being sufficient to bring about the effect, should manifest itself not occasionally, but *whenever there is causation by mental properties*.

As massive overdetermination cannot be the rule, we are led to the conclusion that it is ultimately *in virtue of* some necessarily co-instantiated physical property P that the instantiation of M causes

the instantiation of P^* (and, therefore, also the instantiation of M^*) This means that all causal powers involved in the instantiation of physical properties turn out to be the ones associated to physical properties But if the instantiation of the supervenient property M (*qua* instantiation of M) has no independent causal power to bring about an instantiation of the physical property P^* , it is hard to understand how it could exert any influence on higher-level phenomena as well—no mental causation without mental-to-physical causation

If we use “property-causation” to refer to the relation by which the instantiation of a property X causes an event of type Y *in virtue of* being an instantiation of X (and not in virtue of being an instance of some other co-instantiated property Z), the main steps of the overdetermination argument can be summarized as follows

- (i) M property-causes M^* [higher-level causation]
- (ii) The instantiation of P^* determines the instantiation of M^* [supervenience]
- (iii) M property-causes M^* by property-causing P^* [downward causation]
- (iv) The instantiation of P determines M [supervenience]
- (v) P property-causes P^* [causal completeness of physics]
- (vi) The instantiation of P^* is simultaneously caused by the instantiation of M and the instantiation of P [overdetermination]
- (vii) Overdetermination cannot be the rule
- (viii) M must be identified with P , on pain of being “epiphenomenal”

The overdetermination argument generalizes It does not directly concern mental properties It focuses instead on the relationship between higher-level properties in general—be they mental or not—and those properties defining their supervenience-base As the latter can eventually supervene on more fundamental properties—supervenience is a transitive relation—we may assume that there is a last level made up of absolutely basic properties on which all other properties ultimately depend Basic properties in this sense are conceived of as properly *physical* properties to be identified by the future development of fundamental physics Basic properties are also assumed to

be the only properties connected by genuine laws of nature, so that the principle of causal completeness of physics applies to the physical world as defined by those basic properties. Macrophysical properties as well as functional properties are of course nonbasic: they supervene, as we may suppose, on the properties discoverable by future physics. Thus their claim as real causal factors is also challenged by the overdetermination argument. Specifically mental causation is, therefore, not the main target of the argument. The power of any supervenient property, whether mental or macrophysical, to exert autonomous causal influence on basic phenomena is what is at stake.

3. Supervenient causation

It is of course undeniable that mental properties help explain physical or at least physically constituted phenomena. And explanations based on mental properties are causal. It follows from this that any plausible account of mental properties must give them at least *prima facie* causal efficacy. The normal way to square the attribution of causal explanatoriness to mental properties with the main thrust of the overdetermination argument consists in making whole higher-level causal relations dependent on the causal processes at the basic level. This account, of which there are many variations, came to be known under the title of supervenient causation.

The supervenient causation account says that *A* superveniently causes *B* if *A* supervenes on *A'* and *B* supervenes on *B'* and *A'* causes *B'*. According to a recent attempt to flesh out the idea of supervenient causation (Noordhof 1999), supervenient properties are efficacious because (i) the instantiation of one of their minimal supervenience-bases is a cause of an event of type *E* and (ii) each minimal supervenience-base is such that all its instantiations would cause events of type *E* in some causal circumstances *C*. I will skip here the details of the exact definition of a minimal supervenience-base. For the present purposes, it suffices to observe that minimal supervenience-bases are sets of atomic physical properties. Typically, there will be more than one minimal supervenience-base for any

higher-level property *F*. This is tantamount to saying that the property *F* is multiply realizable. For *F* to be causally efficacious in the process by which another higher-level property *G* is instantiated, two conditions must be satisfied. The first condition says that a minimal supervenience-base of the instantiation of *F* causes a minimal supervenience-base of *G* in the given circumstance. The second condition states that “each minimal supervenience-base of *F* is such that all its instantiations would cause an instantiation of one of the minimal supervenience-bases of *G*, if they were in some causal circumstances *C*—where *C* may vary for each instantiation of *F*” (Noordhof 1999, p. 307). This can be easily applied to the psychological case.

4. Minimal activity?

It remains to be asked whether the supervenient causation account also honor the requirement that causes are, in the actual circumstance, counterfactually necessary for their effects. This is an important requirement. Indeed, causal claims imply certain counterfactual links. Thus in stating that a certain fact causally explains another fact, we assume that the latter fact would not have obtained if the former fact had not obtained. In other words, the truth-conditions of causal claims involve counterfactual conditionals. This means, in a physicalist framework, that causal factors made up of instantiations of physically basic properties must be shown to be causally sufficient *and* counterfactually necessary for any effect that gets produced. As I will presently argue, this is the place where the supervenient causation account founders. The closer to the supposedly basic factors it gets, the less able it is to formulate counterfactually necessary conditions for the effects in question.

Consider the relatively simple case where we invoke the property of being air in a causal explanation of the combustion of a match. This is an explanation of a macrophysical event by means of a macrophysical property. However, macrophysical properties share with mental properties the status of not being physically basic. They are, so we may suppose, supervenient properties whose causal contribu-

tion to their effects should be accounted for in terms of supervenient causation. So we are justified in asking: can we extract, even if only in principle, the ultimate physical causal factors ("the physical cause") from this macrophysical cause?

Presumably, only part of the instantiation of the property of being air—an instantiation of the property of being oxygen—is necessary for the combustion to occur. The property of being nitrogen, for instance, whose instantiation is also part of the instantiation of the property of being air, seems to play no causal role in the events leading to the combustion of the match. The supervenient causation approach is intended to capture cases like this. The property of being air is not competing with the property of being oxygen to play the role of the real physical cause of the combustion. Rather, what we should say here is that the property of being oxygen is the efficacious factor behind the supervenient cause in this case, namely, the presence of air. This is the first move we have to make if we intend to make the distinction between higher-level causes and efficacy determining causal factors coextensive with the distinction between supervenient properties and absolutely basic properties.

However, the same reasoning that leads to the attribution of causal idleness to part of the property of being air can be repeated if we start with the more basic property of being oxygen. After all, not all the oxygen present may be strictly necessary for the combustion to occur. Presumably, part of the oxygen can be discounted as causally superfluous in the process by which the combustion is brought about. What we must recognize here is that being oxygen is one of those properties that admit a certain degree of variation in their "parameters"—in our case, the volume of oxygen available—without prejudice of their role as realizers of higher-level causes. But then a definite value of the relevant parameter of the property of being oxygen will be typically present in a given case in which the presence of oxygen is causally responsible for the combustion. Let us assume that a volume X of oxygen is present in the actual circumstance. Then "having a volume X of oxygen" could be tentatively seen as the ultimate working component behind the cause described at a higher-level by "presence of air." "Having a volume X of oxygen" really looks like a *bona fide* physical property that could be causally responsible, at bot-

tom, for the effects attributable, on a higher-level, to oxygen and, on an even higher-level, to air. However, the instantiation of "having a volume X of oxygen" can hardly be seen as counterfactually necessary for its effects. A lesser quantity Y could also do the causal job at hand, provided Y is greater than some critical value of the relevant parameter of the property of being oxygen. Thus, we face a problem: which of the properties "having a volume X of oxygen," "having a volume Y of oxygen," "having a volume Z of oxygen," is counterfactually necessary for the effect in question? Any instantiation of these properties can realize the higher-level cause. But since none of them is counterfactually linked in the right way to the effect, none of them can be seen as an ultimately efficacy determining property in the physicalist's sense.

Now, if the actual value of the parameter volume cannot be seen as the ultimate working property behind the presence of air, maybe *the actual value of the relevant parameter being greater than a definite critical value* can. Or so we may think. This point of view has the advantage of avoiding the potential multiplicity of causally responsible conditions. It would allow for the formulation of *one* satisfactory counterfactual link. After all, so it could be argued, having at least as much oxygen as the critical value strictly necessary for the combustion is itself a strictly counterfactually necessary condition for causing the combustion in the given circumstance.

This is surely right. But it is not what the physicalist needs. The problem with any property characterizable by reference to a threshold or a critical value is that it is not physical in the physicalist's sense. We may assume that there is a physical matter of fact as to the objective, unique value of the threshold. But a condition expressed by "having at least as much as the critical value strictly necessary for the combustion" can be fulfilled by instantiations of many properties that *eo ipso* have to be considered more basic. "Having a volume X of oxygen greater than the critical value" designates a whole family of properties each of which can realize the property invoked in the specification of the threshold condition. These considerations all go to support the conclusion that counterfactually necessary conditions can only be established by reference to a nonbasic property that *superveniently unifies* physically basic properties.

5. Causal completeness revisited

The argument so far explored the connection between causal efficacy and counterfactual relevance. Accordingly, counterfactual relevance of a property is taken to be a necessary condition for its efficacy: if P is efficacious in the actual circumstance, then it is counterfactually necessary for the effect in question that P is instantiated. What the argument so far shows is that if the instantiation of a property defining a threshold condition should be counted as a cause of an event of type E , then this cause is irreducibly nonbasic, that is, nonphysical by the standards of physicalism.

This result seems to be a straight denial of the causal completeness of physics. It should be clear, however, that the result depends on the consideration of the presence of enough oxygen as a cause of the combustion of the match. On the face of it, the presence of oxygen is a standing state that contributes to the occurrence of the event-effect by playing the role of an enabling condition for the operation of a triggering event-cause like the striking of a match. The relevant question at this juncture is whether, by considering the presence of enough oxygen a cause of combustion, the argument so far unjustifiably and unnecessarily assimilates the role played by causal conditions (states) to the role played by events in the causation of further events.

As a matter of fact, accounts of causation offered by philosophers of mind do not usually distinguish between events and standing states (cf. Steward 1997). They regard events and states alike as particular entities, as causal antecedents which interact on the same footing to bring about a certain effect. This means that enabling conditions and particular occurrences ("concrete individuals") are usually treated as homogeneous factors combining in the causal chain leading to an effect. Indeed, state-like conditions and particular occurrences are treated as *partial causes* that together necessitate the event-effect.

The argument developed in the preceding section follows this common practice in philosophy of mind. It takes the standing instantiation of a property (the property of being enough oxygen) to be the cause of an event of type E . The immediate consequence of showing that this cause is irreducibly nonphysical (in the physical-

ist's sense) is the denial of the causal completeness of physics. This is a high price to pay for claiming autonomy for any special science having to do with supervenient properties.

There is, however, an alternative way to this view, an alternative that could preserve the main point of the argument while still keeping to the causal completeness of physics. It starts with the categorical distinction between events and states, the case for which has been forcefully made by Helen Steward (cf. specially Steward 1997, chapter 7). Accordingly, it is not wrong to insist that a particular event-cause must combine with an independent standing condition to give rise to effects. What is wrong, or at least misleading, is the idea that this is a case of *partial causes* combining in the production of an event-effect. Consider again the example of a match being lighted. For the striking of the match to trigger the desired effect a necessary condition must be satisfied—there must be enough oxygen around. In all nomologically possible worlds in which this condition is not satisfied the particular event-cause is not followed by the lighting of the match. It is misleading to conceive of what is lacking in these worlds as another partial cause, as this may suggest the absence of another particular beyond the triggering factor referred to by “the striking of the match.” Clearly, what has to be given in the actual situation for the particular event-cause in question to bring about the lighting of the match is not a particular entity (which can be represented by a singular term), but a fact (which has to be represented by a sentence).

The crucial point is that being a kind of fact, a standing state bears a relation to the effect it helps to produce which is very different from the relation connecting an event-cause (a particular) and an event-effect (another particular). In Steward's terminology, the first relation is “the relation of causal relevance.” Its expression is a “sentential causal claim.” The second is “the relation of causing,” which is expressed by a “singular causal claim.”

If we now take this into account in the interpretation of the argument developed in the previous section, we arrive at a new result concerning the compatibility of autonomous causation at the higher-level and the physicalistic commitment to a complete specification of the causes of any effect exclusively in physical terms. The remarks in

the previous section draw on the connection between the causal efficacy of properties defining standing conditions for causal processes, on the one hand, and the possibility of establishing counterfactually necessary connections between these properties and the effect in question, on the other hand. In other words, counterfactual significance of a property is used as a test of the existence of a "relation of causal relevance." Some supervenient properties like "presence of enough oxygen" pass the test, while the corresponding physically basic properties in their minimal supervenience-bases do not pass the test. Thus some states can be causally related to physical effects (via counterfactual links) without being themselves physical (in the physicalist's sense).

The existence of a relation of causal relevance connecting non-physical states—which are ontologically kinds of facts, not particulars—with effects implies nothing at all about the possibility of describing the corresponding "relations of causing" exclusively in physical terms. From the point of view of causal relevance of facts or conditions, it is entirely open whether we are able to designate the particulars involved in a causal process in physical terms alone. In other words, "sentential causal claims," as expressions of relations of causal relevance, cannot dictate the form of singular terms in "singular causal claims," which express relations of causing between particular events. In particular, no assertion of a counterfactually grounded relation between an irreducibly nonphysical causal condition and an effect can show the futility of a purported translation of causal claims relating to particular events into the language of fundamental physics. For all we know, this translation may succeed. We have only to keep in mind that this sort of translation does not represent the vindication of reductive physicalism.

We now have the means to formulate the principle of the causal completeness of physics in such a way that it is not contradicted by the main argument of the previous section. The principle says that, for any particular event whatsoever, the chain of previous events connected to it by the relation of causing contains only particulars which can be completely designated by physical terms alone. That these particulars have sometimes to combine with nonphysical facts or conditions to bring about effects does nothing to change their status as physical entities.

References

- Corbí, Josep E and Josep L Prades (2000) *Minds, Causes, and Mechanisms* Oxford Blackwell
- Heil, John and Alfred Mele (eds) (1993) *Mental Causation* Oxford Clarendon Press
- Kim, Jaegwon (1989) "The Myth of Nonreductive Materialism" *Proceedings of the American Philosophical Association* 63 31–47
- (1993a) "Can Supervenience and 'Non-Strict Laws' Save Anomalous Monism?" In Heil and Mele (1993), pp 19–26
- (1993b) "The Non-Reductivist's Troubles with Mental Causation" In Heil and Mele (1993), pp 189–210
- (1998) *Mind in a Physical World An Essay on the Mind-Body Problem and Mental Causation* Cambridge, Mass MIT Press
- Noordhof, Paul (1999) "Causation by Content?" *Mind and Language* 14 291–320
- Steward, Helen (1997) *The Ontology of Mind* Oxford Oxford University Press

Keywords

mental causation, physicalism, causal completeness of physics

Instituto de Filosofia e Ciências Sociais
UFRJ
Largo de São Francisco de Paula 1 Centro
20051-070 Rio de Janeiro RJ
mendonca@ifcs ufjf br