

Questioning the Causal Inheritance Principle*

Ivar HANNIKAINEN

Received: 9.7.2010

Final Version: 16.8.2010

BIBLID [0495-4548 (2010) 25: 69; pp. 261-277]

ABSTRACT: Mental causation, though a forceful intuition embedded in our commonsense psychology, is difficult to square with the rest of commitments of physicalism about the mind. Advocates of mental causation have found solace in the causal inheritance principle, according to which the mental properties of mental states share the causal powers of their physical counterparts. In this paper, I present a variety of counterarguments to causal inheritance and conclude that the conditions for causal inheritance are stricter than what standing versions of said principle imply. In line with this, physicalism may be destined to epiphenomenalism unless multiple realizability turns out false.

Keywords: causal exclusion problem; epiphenomenalism; Jaegwon Kim; mental causation; multiple realizability; Derk Pereboom.

1. *The causal exclusion problem*

It is virtually a platitude of commonsense psychology that our mental states are causally efficacious. Examples abound in daily life which seemingly illustrate that our mental lives are causally relevant with respect to the happenings in the world. For instance, what caused me to go to 7-11 after dinner last Friday (we can call this effect *E*) was my *desire* to buy a pint of Ben & Jerry's paired with my *belief* that 7-11 has a wide selection (let's call this belief-desire pair my mental state *M*). *Voilà*, mental causation. Contrary to the prevailing view in folk psychology, it has seemed far from straightforward to physicalists about the mind how exactly the mental properties of our mental states might introduce their own causal powers in a physical world that is causally closed and complete. The problem of explaining how this might be has been dubbed the *causal exclusion problem* (Kim 1989, 1993, 1998, 2005) and is typically formulated as the inconsistency between some variant of the following four propositions:

1. **Causal Closure of Physics:** If a physical event, *E*, has a cause that occurs at a given time *t*, it has a physical cause, *P*, that occurs at *t*.
2. **Nonreductionism:** Mental properties, *Ms*, are not reducible to, and are not identical with, physical properties, *Ps*.

* A previous version of this paper was presented at the Workshop on "Rationality, Knowledge and Causal Action" at the University of Granada, held on the 17th and 18th of June 2010. Thank you to Helen Beebe, Josep Corbí, María José García Encinas, Christian Piller, Manuel de Pinedo, Murali Ramachandran and Josefa Toribio for their stimulating questions and discussion. Thank you also to the editors of *Theoria* and to an anonymous reviewer for their insightful suggestions and expert guidance.



3. **Mental Causation:** Mental properties have causal efficacy —that is, their instantiations can, and do, cause other properties, both mental and physical to be instantiated.
4. **Exclusion:** No single event, *E*, can have more than one sufficient cause occurring at any given time *t*, unless it is a genuine case of causal overdetermination.

These premises have all seemed plausible to many philosophers in the debate about mental causation but, evidently, one of them must go in order to be left with a coherent picture. The causal closure of physics claims that the physical world as posited by ideal physics is causally and nomologically complete and closed, *i.e.*, for each and every physical event there is a sufficient cause in prior physical events *plus* the laws of physics which govern them. In the 7-11 case, *E* —or my arrival at 7-11 after dinner on Friday night— has a sufficient cause to be found in the previous physical events, let's say, a series of neural firings and muscle contractions, *P*¹. Nonreductionism about the mental is the view that, although mental properties like *M* do supervene on physical properties like *P*, they are not *reducible* to them. My desiring to buy a pint of Ben & Jerry's ice cream supervenes on, but is not reducible to, the neurophysiological state I'm in when I desire to buy a pint.

Supervenience: If any system *s* instantiates a mental property *M*, at *t*, there necessarily exists a physical property *P*, such that *s* instantiates *P* at *t*, and necessarily anything instantiating *P* at any time instantiates *M* at that time.²

Nonreductionism is buttressed most markedly by the widely-accepted phenomenon of *multiple realizability*. In a classic paper, Putnam (1967) revealed just how stringent the empirical commitments of the, then popular, psychophysical type-identity theory are. For every mental state *M*, the identity theorist —Putnam denounced— is committed to specifying a single neurophysiological state, *P*, which those and only those nomologically possible systems capable of instantiating *M* must realize whenever they instantiate *M*. The perceived unlikelihood that type-identity theory could meet this demand meant its early demise. The prevailing view now says that mental property-types are multiply realizable, in the sense that any mental property-type can be realized by a possibly infinite range of different neurophysiological and physical bases since, in different species and structure-types with different biochemical and physical compositions, they are as a matter of fact implemented in dissimilar ways.

Multiple Realizability: For every physical property *P* and mental property *M* which supervenes on *P*, possibly something instantiates *M* but not *P*.

¹ To be sure, neural firings and muscle contraction are events at the *neurophysiological* level. Admittedly, I'm using a vague notion of 'physical' with which intend to designate everything ranging from neurophysiological, through biological and chemical, down to microphysical.

² I'll assume, along with Kim, that supervenience is a necessary condition for physicalism, and that the consequences of supervenience hold by nomological, rather than logical, necessity.

Furthermore, as Segal has noted, multiple realizability is probably the norm between individual cognizers or even cognizers-during-periods. For certain thin mental properties—he uses the example of the pain caused by burning—there may be a single realizer across a whole species, but “talk of species is not appropriate for lots of interesting psychological states. Consider propositional attitudes, for example. It is very unlikely that a given propositional attitude, such as the belief that Barcelona is beautiful, will have the same physical realizations in different people, or, perhaps, in the same individual at different times” (Segal 2009, p. 87). Although Kim has argued that it is compatible with reductionism, according to the prevailing view, multiple realizability is typically recruited in support of nonreductionism about the mental: reducing a given mental property to its neurophysiological and physical realization bases would involve its replacement by a disjunctive, even a wildly or infinitely disjunctive, property. Disjunctive properties of this sort are not scientific kinds and the regularities in which they figure are neither genuine laws nor explanations (Fodor, 1974; Pereboom & Kornblith 1991).

Exclusion is roughly the claim that overdetermination is exceptional rather than pervasive in the physical world. Consequently, mental causation—which if it exists is certainly pervasive—cannot overdetermine its effects. Unlike the victim whose death, E , is genuinely overdetermined by C_1 , one of the arrows to her heart at t , and C_2 , the other such arrow, proponents of exclusion claim that the mental and physical properties cannot both be sufficient and simultaneous causes of any subsequent effects.

So, let’s look at the picture painted by the causal exclusion problem. According to the causal closure of physics, P is causally sufficient for E . If M is not reducible to P , we have two candidates at hand for the role of sufficient cause of E . We might defend that both M and P are causally sufficient and put forth a version of *causal compatibilism* (Horgan 2001, 2007; Pereboom, 2002). Causal compatibilism makes room for psychology as a genuine, explanatory, and autonomous scientific discipline at the cost of denying exclusion and having to explain away worries about causal overdetermination. If one is persuaded by exclusionary reasoning, the choices are to either assert that P is causally sufficient and M is not, thereby forfeiting mental causation and embracing *epiphenomenalism*, or to claim that M is causally sufficient and P is not, in which case we violate the causal closure of physics³. These are dreary and unpalatable outcomes to the robust physicalist who is also a mental realist. If, as Kim has argued, M is reducible to P , which is notoriously a matter of dispute among contemporary physicalists, then by denying nonreductionism we can save mental causation without violating exclusion. On this view, mental causation happens in virtue of the identity or reducibility relation between a mental property and any one of the physical properties which realize it, such that the causal path from M to E coincides with the causal path from P to E . None of the alleged solutions to the causal exclusion problem is altogether satisfying. Each seems to boast some intuitive appeal at the cost of some counterintuitive loss or other: alas, the nature of the beast of causal exclusion.

³ The latter is the implicit solution furthered by counterfactual theoretic account of mental causation, and one that I will argue should not be dismissed too quickly by those seeking to uphold mental causation.

2. Causal inheritance and mental causation

On the physicalist picture of causation introduced earlier and assumed by Kim, Pereboom and by many self-ascribed physicalists, there are no token causal powers other than those occurring at the microphysical level. The only hope for special-scientific causation rests on higher-level properties staking a claim to the causal powers of microphysical properties. In Kim's words:

[T]o be a cause of E , M must somehow ride piggyback on physical cause chains. And we may ask in virtue of what relation it bears to physical property P does M earn its entitlement to a free ride on the causal chain from P to E and to claim this causal chain to be its own (2005, p. 48)?⁴

Reductionists and nonreductionists, of course, give quite different answers to this question and defend corresponding versions of the *causal inheritance principle*. According to reductionists, for higher-level properties to be causally efficacious, they must be reducible to the microphysical level. If they are, then any token causal powers of a higher-level property will be *identical* with some token microphysical causal powers. The reductionist counters the multiple realizability argument by arguing for *local reducibility*. It is an assumption of the defender of multiple realizability, that every realizer P of a mental property M is nomologically sufficient for M . But, as Kim (1989, p. 38) has argued, multiple realizability seems in fact to presuppose laws of a “somewhat stronger form”, according to which, relative to species or structure s_i , a physical state, P_i , is both necessary and sufficient for the occurrence of mental state M ; *i.e.*, $s_i \rightarrow (M \leftrightarrow P_i)$. Thus, the conclusion we ought to draw, according to Kim, is that “the multiple realizability of the mental has no antireductionist implications of great significance; on the contrary, it entails, or at least is consistent with, the local reducibility of psychology, local relative to species or physical structure-types” (1989, p. 39).

Causal Inheritance Principle (CIP): If mental property M is realized in a system at time t in virtue of physical realization base P , the causal powers of *this instance of* M are identical with the causal powers of P (Kim 1993, p. 326).

By contrast, nonreductionists, such as Horgan and Pereboom, argue that the weaker metaphysical relation that according to them holds between mental and physical properties and warrants the irreducibility of the former to the latter is, nonetheless, strong enough for them to share causal powers. Unlike Kim, they take multiple realizability—as is standard practice—to support nonreductionism. In addition, because M is multiply realizable, not only is M token distinct from its physical realizer P , but M 's causal powers must also be distinct from, but “nothing over and above”, those of P . Thus,

Weaker Causal Inheritance Principle (WCIP): If mental property M is realized in a system at t in virtue of physical realization base P , the

⁴ This quote has insignificant relettering to cohere with my choice of M for ‘mental property’, P for ‘physical property’ and E for ‘effect’. In the remainder of this paper, I take the liberty to make similar, minimal changes to a quote where the author has preferred to use other letters.

causal powers of this instance of M are wholly constituted by the causal powers of P (Pereboom 2002, p. 504).

Causal inheritance is an intuitive and plausible way to sneak mental causation into a physicalist picture of causation at large. It may, moreover, be a crucial element of any successful and persuasive defense of mental causation. My contention, however, is that the causal inheritance principle has not been carefully formulated and is subject to a variety of counterarguments against which it should be defended by physicalist advocates of mental causation. More careful philosophical attention ought to be placed on the phenomenon of causal inheritance and, consequently, on a precise and correct formulation of the principle. Throughout Sections 3 to 5, I will argue that the conditions under which causal inheritance takes place are stricter than what the standing versions of the causal inheritance principle imply.

3. Is the causal inheritance principle valid?

The causal inheritance principle purports to specify the conditions under which, and the extent to which, an instance of a mental property can be said to inherit the causal powers of its physical realization base⁵. In the CIP formulation, it suffices for M to be *realized or instantiated in virtue of* property P . Note that that is a substantially weaker requirement than:

CIP-R: If mental property M is reducible in a system s_i at time t to its physical realization base P_i , the causal powers of this instance of M are identical with the causal powers of P_i .

It seems that CIP-R does have some *prima facie* plausibility (which it will be the object of Section 5 to examine). After all, since for M to be reducible in s_k to P_k entails some identity-like metaphysical relation, it is plausible that, when they stand in said relation, M and P_k confer an identical set of dispositions to s_k and that M thereby inherits the causal powers of P_k . It is likely that this is what Kim indeed means by CIP and that he simply assumes that for M to be realized by P *just is* for M to be reducible to P . Nevertheless, I think this point is noteworthy because, so formulated, CIP is compatible with supervenient causation. Suppose it turned out that nonreductionism were the truth about the mind-body problem, *viz.*, that M supervenes on, but is not reducible to, P . On this picture, M would indeed be realized by P and would, by CIP, inherit the causal powers of P . But Kim explicitly shuns this kind of causation:

[...] why not say that M , though it doesn't quite have the status of P in relation to E , is a derivative cause of E in virtue of its supervenience on P ? M is not an independent cause of E ; its causal status derives from its supervenience on the causally active Pinserting

⁵ For a substantive account of causal powers, see Mumford, 2009. Kim doesn't say much about what he takes causal powers to be, so it'll be helpful to keep Mumford's account in mind. In a nutshell, Mumford defends (i) that "powers are the causes of their manifestations", that (ii) events are *polygenic* in that "they are produced by many powers working together, or against each other," and that (iii) powers are *pleiotropic*, i.e., they "can contribute to many different types of event in different circumstances" (Mumford 2009, pp. 102-4).

a broken-dot arrow [between M and E] and calling it supervenient causation, or anything else (how about “pretend” or “faux” causation?), does not alter the situation one bit. It neither adds any new facts nor reveals any hitherto unnoticed relationships (Kim 2003, p. 171).

It seems that Kim would want his formulation of the causal inheritance principle to be foolproof against this possibility. Since, even if he doesn't, Kim's account seems to demand reduction as a condition for causal inheritance, I will proceed in Section 5 as if CIP-R is the intended formulation of the causal inheritance principle for conservative reductionism and deal in Section 4 with causal inheritance in a nonreductionist framework.

4. Nonreductionism and causal inheritance

How can irreducible mental properties inherit the causal powers of their physical realizers in virtue of the former's mere supervenience on the latter? There are two versions of this question corresponding to two different degrees to which causal inheritance is defended by nonreductionists. Horgan, in his causal contextualist account (2001), seems to assume that *full causal inheritance* is warranted between supervenient and subvenient properties in virtue of something like CIP; while Pereboom's causal compatibilism (2002) makes the weaker claim, which I have labeled *partial causal inheritance*, that supervenient mental properties share *some* of the causal powers of their physical realizers in virtue of WCIP. The object of this section will be to tackle these two varieties of causal inheritance in a nonreductionist framework and ultimately to argue that neither CIP nor WCIP hold for irreducible mental properties.

4.1. Full causal inheritance: the case of causal contextualism

In his classic “Mental Causation”, Yablo brought back the determinate-determinable relation to characterize the psychophysical, arguing that “such a view is in fact implicit in the reigning orthodoxy about mind-body relations, namely, that the mental is *supervenient* on, but *multiply realizable* in, the physical” (1992, p. 254). On his view of properties,

P determines Q ($P > Q$) only if:

- (i) necessarily, for all x , if x has P then x has Q ; and,
- (ii) possibly, for some x , x has Q but lacks P (Yablo 1992, p. 252).

In other words, when M supervenes on and is multiply realizable by $P_1, P_2 \dots P_i$, then $P_1, P_2 \dots P_i$ can be said to determine M . Similar pairs of *determinable* and *determinate* properties are *being red* and *being scarlet*, *having a shape* and *being triangular*, *having a weight* and *weighing three kilograms*, etc. Yablo appealed to this determination relation ultimately to shed light on how M might outdo P in the competition for causal relevance in virtue of its proportionality to E . Here I intend to apply this relation for other purposes: namely, to show how, for some class of E s, E causally depends on P but not M .

Suppose I drive home and thereby instantiate the property of *speeding home*. Since *speeding home* is *driving home* in a specific way, *i.e.*, a determinate of *driving home*, whatever

instantiates the *speeding home* property instantiates also the *driving home* property. Now we may ask: does this instance of *driving home* have the causal powers of *speeding home* as CIP supposes? It seems obvious to me that, on the contrary, the *speeding home* has a range of causal powers —e.g., to make the car skid off an icy road— which the *driving home* does not have. These causal powers are unique to the determinate property precisely because the determinate property is a more refined property than is the determinable property⁶. These are causal powers that other determinations of the relevant determinable property do not share with the determinate property in question; and *eo ipso* causal powers that the determinable property does not share with the determinate property. Thus, it is possible for the subvening or determinate property to have causal powers that the supervening or determinable property does not inherit.

It may be objected that, while the *driving home*-type indeed does not inherit the causal powers of the *speeding home*-type, this instance of the *driving home* property does. This line of response seems to confuse this instance of a driving home event, which has the causal power to make the car skid off an icy road (in virtue of the causally relevant property which we are trying to identify), with this instance of a *driving home property* which does not have such causal power.

Compare the following two causal explanations identifying the causally relevant property. The causal explanation:

My car skid off the icy road because I was driving home

sounds like it is missing crucial, causally relevant information which

My car skid off the icy road because I was speeding home

is not. The paucity of supervenient causal explanation stems from the fact that it leaves out information about the subvening property which is crucial to the event's causal powers. Invoking the *driving home* property in the first causal explanation to explain what made my car to skid off the icy road is a way of leaving out that kind of causally relevant information. By contrast, citing the *speeding home* property cited as the relevant property in virtue of which the driving home event led to the car's skidding off the icy road results in a robust causal explanation. It, of course, won't help advocates of supervenient causation to insist that the *driving home* property inherits the causal powers of the *speeding home* property by saying that this instance of the *driving home* property in fact *is* a *speeding home* property: this is reductionist, not supervenient, causal inheritance.

It may be helpful to support the above case with an example from the psychological domain. Consider a studying event which results in my burning 50 extra calories. What was the causally relevant property in virtue of which my studying resulted in my *burning 50 extra calories*? Was it *M* or *P_k*, i.e., the supervenient property of *memorizing every word on the GRE Hot List* or some subvenient neurophysical property *involving in-*

⁶ The more refined property, on Yablo's view I am adopting here, is the property with the larger cumulative essence or the property it is "more difficult to be" insofar as "a thing's [cumulative] essence specifies what it takes to be that thing" (Yablo 1987, p.299).

tensive neural communication between the hippocampus, medial temporal lobe and other memory-related areas? It seems here too that P_k is what caused me to burn 50 extra calories. If I were a MacBook Pro, a Leibnizian monad or simply another human being with a slightly different brain, blood glucose level or heart rate, I would instantiate property M by instantiating some different physical realizer P_l and without causing precisely the same effect. Contrarily, it seems like someone exhibiting the very same, complex subvenient property, P_k , would also thereby burn 50 extra calories in all worlds compatible with the laws of physics. Thus, certain fine-grained, *realization-dependent* (RD) effects are caused by P_k and it is strictly speaking wrong to insist that, by causal inheritance, M is the cause of those effects too.

Contra full causal inheritance, tokens of multiply realizable, supervenient properties do not share the causal powers of their subvenient properties; and, *a fortiori*, mental properties do not inherit all the causal powers of their physical realization bases. This conclusion has obvious ramifications for Horgan's (2001) compatibilist defense of mental causation according to which causation is a "contextually-parameterized notion". According to Horgan (2001, p. 102):

A single phenomenon can perfectly well be subject to a variety of different causal explanations, involving properties from a variety of different counterfactual-dependence patterns at different descriptive/ontological levels. [...] The compatibility of these different levels of explanation largely stems from inter-level supervenience relations. Since the higher-order, psychological patterns and generalizations are supervenient upon underlying physical facts and laws, the mental properties that are causal properties at the psychological level have their causal efficacy via the causal efficacy of physical causal properties that realize them.

The problem with this view is that, for any RD effect, E_{rd} , the mentalistic counterfactual-dependence pattern does not hold. If you will, the counterfactual $P \rightarrow E_{rd}$ is true while the counterfactual $M \rightarrow E_{rd}$ is false. Consequently, for any RD effect, there is no true mentalistic counterpart to the physical counterfactual dependence pattern for the contextual parameter to choose from.

4.2. *Partial causal inheritance: the case of causal compatibilism*

The above conclusion is one that nonreductionists such as Pereboom who endorse WCIP are willing to admit. Pereboom (2002, pp. 503-4) agrees that if nonreductionism turned out to be true:

[...] this reflection would also undermine a token-identity claim for mental causal powers—should they exist—and their underlying microphysical causal powers. For if the token microphysical realization of M had been different, its token microphysical causal powers would also have been different. We therefore have good reason to suppose that any token mental causal powers of M would not be identical with the token microphysical causal powers of its realization.

Still, there would be a sense in which the token causal powers of M would be 'nothing over and above' the token causal powers of P [...] For this reason, it makes sense to say that token mental causal powers are wholly constituted by token microphysical causal powers.

The central, and admittedly more plausible, claim of partial causal inheritance accounts is that there is a class of causal powers of mental events that are *realization-independent* (RI), which M can be rightfully said to share with P^7 . In contrast with the RD effect considered above, we might for instance say that the RI effect of *acing the GRE Verbal section* is one that M does cause. Here it seems that the details of the physical realization base P_k of M are irrelevant to the manifestation of the effect (or, in Yablo's terms, that the "effect is relatively insensitive to the finer details of the physical implementation" (1992, p. 188)). That is, as long as I instantiated M , this RI effect would have ensued, whether I were a Mac Book Pro, a Leibnizian monad or a studious and gifted college senior. Pereboom's argument in defense of partial causal inheritance, if I understand it correctly, makes this sort of claim.

One way to back up this claim, of course, is to follow Yablo's (1992) lead in showing how modal discernibility between M and P_k renders M a *better* (*viz.*, more proportional) candidate for the role of cause of those RI effects than is P_k . However, Yablo's approach makes use of a counterfactual account of causation-as-dependence that differs markedly from the physicalist account of causation-as-production presupposed in the causal exclusion problem. Thus, Pereboom is wise not to consider this option, of using modal discernibility of M and P in his argument for mental causation, if he is to consider his a robust, nonreductive *materialism*. Instead, he embraces the physicalist view of causation according to which microphysical properties are inherently causally efficacious and then argues that, owing to WCIP, M inherits some causal powers from its physical realization base, P . The crux of Pereboom's view is that, contrary to the orthodox physicalist claim that "the causal powers that have a role in explaining the dispositional features of mental states are nondispositional properties of their realization bases", a robust nonreductionist should "consider the possibility that there are properties intrinsic to mental state types that play a part in explaining their dispositional features" (2002, p. 516). Consequently, he believes that the causal generalization

m 's instantiating M causes E_{π}

is *bona fide* and indispensable (hence, *robust* nonreductionism), and in some substantive sense independent of the causal generalization

m 's instantiating P_i causes E_{π} .

This move, of making M intrinsic to m , is supposed to safeguard his account from the worries about causal inefficacy with which *relational* (functional) mental properties are riddled. Since functional mental properties are dispositional, they are defined in terms of their RI effects; *i.e.*, M : the property which *ceteris paribus* causes E_{π} .⁸ According to Block (1990), Rupert (2006) and others, this feature of functional properties

⁷ This strategy is at the heart of Yablo's (1999) defense of mental causation too. RI effects are, in Yablo's terminology, those to which M is *proportional*.

⁸ Previous critics typically call these effects the 'characteristic' effects of functional properties. As far as I can tell, the characteristic effects of a functional mental property *are* the realization independent effects, *i.e.*, the effects that a functional mental property has independently of the details of the realization base by which it is instantiated.

renders causal generalizations in which the instantiation of a functional mental property figures as one of the causal relata, *i.e.*, generalizations of the form

m's instantiating *M*, the property which *ceteris paribus* causes E_{α} , causes E_{β}

redundant in a way that undermines their explanatory power. These explanations express logical rather than nomological relations between a functional property instantiation and an RI effect and so are not causal explanations by any scientific realist standards. In an attempt to defend mental causation in a nonreductionism framework, Pereboom eschews functionalism by making the causally relevant mental properties *intrinsic* to mental state types. Let us call this Pereboom's robust nonreductive dictum (RN).

RN: A mental state *m* type has an intrinsic mental property *M* that explains some of the dispositional features, *e.g.*, the causal powers to bring about RI effects, of an instance of *m*.

In the remainder of this section, I will argue that this move doesn't get very far in making the mental property causally efficacious with respect to its characteristic RI effects. Not unlike in the problem of metaphysically necessary effects outlined above, mentalistic regularities according to RN are shown not to be properly causal explanatory.

Since *M* is intrinsic to *m* according to RN, every instance of *m* has *M* regardless of what relations it enters into with the world around it. Additionally, on any view of events as property instantiations, *M* is *essential* to *m*. An obvious consequence of this is that RN makes *M* a categorical property of the *m*-type, *i.e.*, a property that *m* has everywhere that it exists and regardless of the relations that *m* enters into with the world around it. This is at least consistent with the criticism directed at functionalism that at least some, phenomenal mental properties like the felt experience of *painfulness* are archetypal categorical properties (Yablo 1987, p. 313). But this, as the remainder of my argument will attempt to elucidate, is an important hindrance for the RN view because the categorical properties of an event-type do not stand in causal relations to the event-type's characteristic dispositions, among which are its powers to bring about RI effects.

Consider the following example from Mumford about the disposition *D* of a billiard ball to "roll in a straight line on a flat surface when struck":

Various things can roll, but what something must possess if it is to have the disposition is a surface, various points of which come into contact, at different times, with the surface upon which it rests, so that a change in location is possible. The disposition to roll an indefinite distance, under suitable circumstances, is possessed by both wheels and billiard balls; all that is necessary for this disposition is the possession of a single circumference which is rigid to a certain degree. [...] A billiard ball can do something a wheel cannot, however: it can roll instantly in any direction, according to where it is struck, whereas a wheel is limited to certain directions[...]. (Mumford 1994, p. 428).

So, the billiard ball also possesses categorically "an infinite number of circumferences"; also, since it rolls in a straight line, "it is composed of a substance of even mass throughout", and so on. Thus, we arrive by empirical investigation at some cate-

gorical property or property complex C , *being spherical*, which grounds or “supports” the dispositional property in question, D , *rolling uniformly in a straight line when struck*. We must agree with Mumford that:

The categorical property C to which D is correlated is indeed correlated contingently, but only insofar as the laws of physics are contingent. Given certain laws of nature, the correlation of C to D is necessitated (Mumford 1994, p. 429).

So the problem with RN is not, like with functionalism, that mentalistic causal generalizations express logical relations —because they, in fact, express nomological relations— but rather that they don’t express the right *kind* of nomological relation, *i.e.*, they express what I am calling *expository* rather than causal relations. On this view, we cannot say that the billiard ball’s *being spherical*, an archetypal categorical property, caused it to *roll uniformly in a straight line when struck*, an archetypal dispositional property. By this same reflection, the categorical mental property or property complex of a mental event does not cause the dispositional properties, among which figure its powers to bring about RI effects.

To see this, consider another set of regularities that are logically contingent but nomologically necessary, *viz.*, criminal laws, and specifically the law regarding underage drinking. (In order to draw out the parallel between legal and causal regularities, I will pretend that *legal* laws express “causal*” relations.) To draw the analogy with RN, we need to assume that *being underage* refers to a categorical property of s_k that supervenes on, but is not identical to, s_k ’s *being n_k years old* such that anyone who is n_k years old is underage. Firstly, *being underage* is essential to s_k , since s_k cannot be an s (a minor under a specific legislation) without having the *being underage* property, while *being n_k years old* is accidental, because s_k could have some other specific age, *e.g.*, *being n_l years old*, and still be a minor. Secondly, both *being n_k years old* and *being underage* are intrinsic properties⁹. Now consider the RI effect* of *being denied the purchase of alcohol*. It is RI because, regardless of an s ’s specific age, as long as the minor’s being her actual age entails her being underage, she will be denied the purchase of alcohol. We have two potential causal* explanations for this RI effect:

Janet was denied the purchase of alcohol at t because she was underage at t ,
 Janet was denied the purchase of alcohol at t because she was n_j years old at t ,

⁹ One might sensibly object that *being underage* is extrinsic since whether s_k is underage depends on what particular age is stipulated by the relevant authority as the upper bound of minority. However, as is widely accepted, whether a property appears intrinsic or extrinsic is a matter of what we hold fixed. The speed of light, c , is by usual standards an intrinsic property because we assume, as part of the invariable background conditions, the set of actual physical laws which guarantees that the speed of light is c . If we considered the suggestion that there are possible worlds with different sets of physical laws, then it would seem that c is an extrinsic property of light which depends on the set of physical laws in place. So, for the proposed analogy between legislation and physical laws to hold, let us hold fixed a *single* legislative code such that, if s_k is underage, any intrinsic duplicate of s_k situated in a different context will also be underage. This has the effect of making the *being underage* property intrinsic (to all minors under the relevant legislation) just as *mass* or *speed* properties are intrinsic (to physical objects in all worlds compatible with the actual physical laws). I owe thanks to an anonymous reviewer for requesting this explanation.

and two potential causal* generalizations (laws) to account for the above case:

ELUD: s 's being underage at t causes* s 's being denied the purchase of alcohol at t .

CLUD: s 's being $\leq n_i$ years old at t causes* s 's being denied the purchase of alcohol at t .

The first generalization, ELUD, is *expository*. It fleshes out what minority is by law, by relating two properties that all minors share, *viz.*, the dispositional property of *being denied the purchase of alcohol* and the categorical property of *being underage* that nomologically supports the former. CLUD, on the contrary, has the explanatory character of a *causal** generalization. By elucidating the range of ages, 0 to n_i , that result in being denied the purchase of alcohol, CLUD informs us about the nomological regularities in place in Janet's country, *viz.*, the law concerning underage drinking. In this instance of causation*, to say that "Janet was denied the purchase of alcohol at t because she was n_j years old at t ", where n_j is an age such that anyone who is n_j years old is underage, seems to be a stronger candidate for a causal* explanation while "Janet was denied the purchase of alcohol at t because she was underage at t ", is more of an expository claim.

On RN (as on the functionalist view), mentalistic causal explanations —"Tim cried because he was sad", "Vanessa went to the ice cream stand because she was craving ice cream"— fail, in my ears, to do any serious explanatory work. They seem decidedly in the business of making expository claims about categorical properties and the dispositions they support by nomological necessity. Thus, the relation between *being sad* and *crying* is of the same kind as the relation between *being underage* and *being denied the purchase of alcohol* or between *being spherical* and *rolling uniformly in a straight line when struck*: a relation of nomological support and not of causation¹⁰. *Pace* RN, in order to genuinely *explain* the disposition of a mental state m to bring about E_{ri} and thereby discover a nomological regularity in place in the actual world, one must replace the categorical supervenient property, M , with the range of accidental subvenient properties which realize M by necessity. Thus, the principle by which M inherits the causal powers of its physical realization base, WCIP, is invalid: if RN is true and M is categorical to m , M supports, rather than causes, m 's dispositional properties among which are its powers to bring about RI effects.

5. Reductionism and causal inheritance

Contemporary reductionist accounts of the mind typically admit multiple realizability and, thereby, defend that mental properties are identical to the disjunction of their physical realizers, actual and possible. In this section, I will try to show that it is difficult to see how a disjunctive (mental) property, even if we make room for it in our ontology, can inherit the causal powers of its disjunct (physical) properties via CIP-R.

¹⁰ See also Segal's example about the "non-strict law that objects with that very specific sort of red glow [that red-hot poker get] cause wax to melt" (2009, p.91).

It is perfectly meaningful to say that “Wine is either red or white” and, thus, unproblematic to apply the predicate “being red or white” to wine. It seems clear that we do not need to postulate the existence of the *being red or white* property to explain why a disjunctive predication is true. In the arena of psychophysical reduction, this counts as a reason to endorse an *eliminative* reduction about mental properties. This is not a promising avenue toward mental causation via causal inheritance since the disjunctive predication “being M ” fails to correspond to a mental property and refers merely to the disjunction of causally efficacious physical properties. It is a system’s instantiating P_k that explains why it is true to predicate “is P_k ”, and, on the eliminative view, it is also because the system instantiates P_k that it is true of it to predicate “is M ”. In this way the causal explanations, “ M caused s_k to E ” and “ P_k caused s_k to E ” are both true in virtue of s_k instantiating the causally efficacious property P_k which has the causal power to bring about E . This is why, in a nutshell, eliminative reductions are no path to mental causation. More interestingly, we should wonder how mental properties, conservatively reduced, could indeed share the causal powers of their physical realizers via CIP-R.

This consideration requires *ex hypothesi* that we grant the existence of disjunctive properties. An obvious consequence of the existence of disjunctive properties is that, for any system s_i that instantiates a disjunctive property Q , s_i necessarily instantiates also a disjunct property R_i to which Q is disjunctively identical. Of course if R_i is itself disjunctive, then it must also instantiate a disjunct property S_i to which R_i is disjunctively identical and so on. Ultimately, any system that instantiates a disjunctive property, must also instantiate some sparse, physical property. In the context of mind-body relations, we are left with the disjunctive identity relation between a mental property and its diverse physical realizers, actual and possible:

$$M = P_1 \vee P_2 \vee P_3 \dots \vee P_n.$$

Now consider the disjunctive property,

$$\textit{being wine-colored} =_{\text{df}} \textit{being red} \vee \textit{being white}$$

On the conservative reductionist view, if “The wine bottle I am about to open is wine-colored” is true, it must be the case that the relevant wine bottle has the *being wine-colored* property (and, additionally, has either the *being red* or the *being white* property). What we cannot reasonably defend is that “The wine bottle I am about to open is wine-colored” is true *in virtue of* the bottle’s *being wine-colored*. The truthmaker for the above predication is some sparse property of the wine bottle, *i.e.*, the wine bottle’s being a particular shade of red or white (Crane, 2008). Similar reasoning renders mentalistic causal explanations true in virtue of causally efficacious physical properties. To see this, suspend any disbelief about color psychology and suppose that, according to recent findings in color psychology, there is a detailed account of how looking at a sample of red causes the observer to feel warmth and comfort and looking at a sample of white causes the observer to feel purity and peace. Now consider the causal explanations with the disjunctive property *being wine-colored*.

Hannah felt warmth and comfort because she contemplated this wine-colored wine bottle.

James felt purity and peace because he contemplated that wine-colored wine bottle.

We can, unproblematically, concede that these causal explanations are true. Nevertheless, the truthmakers for the above causal explanations in each case are the corresponding causal generalizations below:

CP1: *s*'s contemplating red causes *s*'s feeling warmth and comfort

CP2: *s*'s contemplating white causes *s*'s feeling peace and purity

So far, so good: an advocate of CIP-R can agree to this while insisting that we should think of an instance of the *being wine-colored* property as inheriting the causal powers of the disjunct property which realizes it at a given time. Well, what if we did as she recommends? It is easy to see that the kind of causation that *being wine-colored* would enjoy if we granted its causal inheritance would be a lawless, non-generalizable kind. There would be no predicting what a *wine-colored* thing would do *qua wine-colored*. (This is guaranteed by the multiple realizability of *being wine-colored*, *i.e.*, the fact that the disjunct properties to which *being wine-colored* is disjunctively identical are not identical amongst themselves.) The causal powers of a wine-colored thing would be, in some structure-type *s*, to give rise to feelings of warmth and comfort and, alternatively, in some different structure-type *s'*, to give rise to feelings of peace and purity, and in some other possible structure-type, a different range of causal powers yet! If you will, RD effects crop up all over again in a reductionist framework. Keep in mind, though, that the causal powers of disjunctive properties would not be genuinely anomalous, since, at the level of their disjunct physical realizers, they would be perfectly regular and lawful. In sum, I frankly don't see the benefit of claiming causal inheritance in order to give rise to a kind of *faux*-anomalous causation at the level of disjunctive properties which is itself fully explainable in terms of regular, causal generalizations at the level of their disjunct physical realizers.

A lingering worry might be that —now back to psychophysical reduction— mentalistic causal explanations turn out false if we give up CIP-R. As I have been exhibiting in this section, mentalistic causal explanations remain true even if we surrender to epiphenomenalism about mental properties; and this is because, just as disjunctive predications are true even if we do away with disjunctive properties, mentalistic causal explanations can be true even if mental properties are causally inefficacious: they are true, yes, but their truthmakers are causal generalizations at the level of their physical realization bases.

The arguments I have run through in this paper do not preclude the possibility that, if multiple realizability turned out false, mental properties could be said to inherit the causal powers of their properly identical physical realizers. Consider the *causal inheritance principle for properly reducible properties*:

CIP-R*: If mental property M is properly (i.e., nondisjunctively) reducible in a system s_i at time t to its physical realization base P , the causal powers of this instance of M are identical with the causal powers of P .

If multiple realizability were false and $M_i = P_i$, then the causal powers of M_i would be identical to—and as lawful and generalizable as—the causal powers of P_i in virtue of CIP-R*. Moreover, mentalistic causal explanations would be *bona fide* rather than made true by lower-level physical causal generalizations.

6. Conclusion

The inheritance of causal powers by a multiply realizable, supervenient mental property, M , from its physical realizer, P , is problematic in ways that physicalist advocates of mental causation have ignored. In the case of nonreductionists, this is because it is simply wrong to think that M inherits the causal powers to bring about RD effects from P , since it is possible for M to be realized in some different system by some different physical property and in such cases for the instantiation of M to fail to bring about the same RD effect. *Contra* WCIP, M does not inherit the causal powers to bring about RI effects either, since, if (owing to RN) M is the categorical property of m , then M nomologically supports, rather than causes any dispositional properties of m among which are the powers to bring about RI effects. In the case of reductionism, claiming causal inheritance via CIP-R results in an *ad hoc* variety of lawless, causal powers at the level of disjunctive properties which are fully explainable by the regular, lower-level causal generalizations featuring the physical realizers to which the disjunctive property in question is disjunctively identical. The difficulties with causal inheritance which I have elaborated are silent about the possibility of *intrinsic* mental causal powers. But advocating intrinsic mental causal powers is no easy task for contemporary physicalists. The intrinsic causal powers of mental properties, if there were any, could not make a difference to the physical properties of their effects—because of Causal Closure, *i.e.*, because prior physical properties are already causally sufficient for them—nor could they make a difference to the mental properties of their effects—because of Supervenience, *i.e.*, because mental properties supervene on physical properties which, because of Causal Closure, are themselves already causally explained by the physical properties of prior events.

The two options that, I think, remain open to physicalists about the mind are both scientifically respectable. On the one hand, Bickle (2003) has argued that—while, at the systems level, multiple realizability obviously holds—neuroscientific research has found, at the level of cellular physiology and molecular neurobiology, unitary realizers for psychological kinds which could serve as reduction bases and, consequently, has elicited an empirically grounded hope that type-identity theory could resurge. In this eventual case, I have argued, mental properties would share the causal powers of their physical realizers at the cellular and molecular level in virtue of CIP-R*. On the other hand, recent studies in cognitive neuroscience draw out the diverse, unconscious heuristics (Wegner, 2003; Wegner & Wheatley 1999) and illusory mechanisms (Tsakiris & Haggard 2003; Wohlschlagel et al., 2003) at work in subserving the intuition of agent

causation. As this line of research develops, empirically-minded philosophers might be prepared to relinquish the forceful intuition of mental causation in favor of an epiphenomenalism about mental properties —a view which, to be sure, commits one neither to the causal inefficacy of mental *states* nor to the falsity of psychological causal explanations. For those intent on saving mental causation, a third option might be to embrace a counterfactual account of causation *à la* Loewer (2001) and Yablo (1992). Undoubtedly there are numerous obstacles to a counterfactual account; to name a few, Kim has objected that they don't distinguish epiphenomenal relations from genuine causation, and it is unclear whether a counterfactual account, which means abandoning the causal closure of physics, is available to a serious physicalist. If these difficulties are insurmountable, Kim's (2003, p. 165) motto will ring true again, "Either reduction or causal impotence", in which case I hope to have shown that the kind of reduction required to avoid epiphenomenalism is stricter than what Kim himself has in mind.

REFERENCES

- Bickle, J., 2003. *Philosophy and neuroscience: a ruthlessly reductive account*. Dordrecht: Kluwer.
- Block, N., 1990. Can the mind change the world? In G.S. Boolos, ed. *Meaning and method: essays in honor of Hilary Putnam*. Cambridge: Cambridge University Press, pp. 137-70.
- Crane, T., 2008. Causation and determinable properties: on the efficacy of colour, shape, and size. In J. Hohwy, & J. Kallestrup, eds. *Being reduced: new essays on reduction, explanation, and causation*. Oxford: Oxford University Press, pp. 176-95.
- Fodor, J., 1974. Special sciences (or: the disunity of science as a working hypothesis). *Synthese* 28(2), pp. 97-115.
- Horgan, T., 2001. Causal compatibilism and the exclusion problem. *Theoria*, 16(40), pp. 95-116.
- Horgan, T., 2007. Mental causation and the agent-exclusion problem. *Erkenntnis*, 67(2), pp. 183-200.
- Kim, J., 1989. The myth of nonreductive materialism. *Proceedings and Addresses of the American Philosophical Association*, 63(3), pp. 31-47.
- Kim, J., 1993. *Supervenience and mind: selected philosophical essays*. Cambridge: Cambridge University Press.
- Kim, J., 1998. *Mind in a physical world: an essay on the mind-body problem and mental causation*. Cambridge (MA): MIT Press.
- Kim, J., 2003. Blocking causal drainage and other maintenance chores with mental causation. *Philosophy and Phenomenological Research*, 67(1), pp. 151-76.
- Kim, J., 2005. *Physicalism or something near enough*. Princeton: Princeton University Press.
- Loewer, B., 2001. Review of *Mind in a physical world*. *Journal of Philosophy*, 98(6), pp. 315-24.
- Mumford, S., 1994. Dispositions, supervenience and reduction. *The Philosophical Quarterly*, 44(177), pp. 419-38.
- Mumford, S., 2009. Passing powers around. *The Monist*, 92(1), pp. 94-111.
- Pereboom, D., 2002. Robust nonreductive materialism. *Journal of Philosophy*, 99(10), pp. 499-531.
- Pereboom, D. and Kornblith, H., 1991. The metaphysics of irreducibility. *Philosophical Studies*, 63(2), 125-45.
- Putnam, H., 1967. Psychological predicates. In W.H. Capitan & D.D. Merrill, eds. *Art, Mind, and Religion*. Pittsburgh: University of Pittsburgh Press, pp. 37-48.

- Rupert, R. D., 2006. Functionalism, mental causation, and the problem of metaphysically necessary effects. *Noûs*, 40(2), pp. 256-83.
- Segal, G., 2009. The causal inefficacy of content. *Mind & Language*, 24(1), pp. 80-102.
- Tsakiris, M. and Haggard, P., 2003. Awareness of somatic events associated with a voluntary action. *Experimental Brain Research*, 149(4), pp. 439-46.
- Wegner, D.M., 2003. The mind's best trick: how we experience conscious will. *Trends in Cognitive Sciences*, 7(2), pp. 65-9.
- Wegner, D.M. and Wheatley, T., 1999. Apparent mental causation: sources of the experience of will. *American Psychologist*, 54(7), pp. 480-92.
- Wohlschlagel, A. Haggard, P. Gesierich, B. and Prinz, W., 2003. The perceived onset time of self- and other-generated actions. *Psychological Science*, 14(6), pp. 586-91.
- Yablo, S., 1987. Identity, essence, and indiscernibility. *The Journal of Philosophy*, 84(6), pp. 293-314.
- Yablo, S., 1992. Mental causation. *The Philosophical Review*, 101(2), pp. 245-80.

Ivar HANNIKAINEN is a graduate student in philosophy at Universidad Complutense de Madrid. His research interests range from ethics to philosophy of psychology and of language. Concerned with the relationship between the empirical sciences and philosophy, in his dissertation he aims to discover some implications of the psychology of moral agency for normative ethical theory.

ADDRESS: Departamento de Lógica, Lenguaje y Ciencia, Facultad de Filosofía, Universidad Complutense de Madrid, Ciudad Universitaria, 28040 Madrid (Spain). E-mail: ivar.hannikainen@gmail.com