

The Causal Exclusion Argument

Jesper Kallestrup. Forthcoming in *Philosophical Studies*.

Jaegwon Kim's causal exclusion argument says that if all physical effects have sufficient physical causes, and no physical effects are caused twice over by distinct physical and mental causes, there cannot be any irreducible mental causes. In addition, Kim has argued that the nonreductive physicalist must give up completeness, and embrace the possibility of downward causation. This paper argues first that this extra argument relies on a principle of property individuation, which the nonreductive physicalist need not accept, and second that once we get clear on overdetermination, there is a way to reject the exclusion principle upon which the causal exclusion argument depends, but third that this should not lead to the belief that mental causation is easily accounted for in terms of counterfactual dependencies.

I. Nonreductive Physicalism

Physicalism is the view that the world is the way it is, because the physical world is the way it is. All the facts, including all the mental facts, are fixed by the physical facts. Restricting our attention to the mental and physical realms, this means that properties within the former must at least supervene on properties within the latter:

(Supervenience) Necessarily if something has any mental property M, there is a physical property P such that the thing has P, and necessarily anything with P has M;

where the notion of necessity in play is taken to be metaphysical necessity.¹ There are three important points. First, this rules out substance dualism if we assume that mental substances are those with only mental properties. (Supervenience) says that anything with a mental property necessarily has a physical base property, so nothing has only mental properties. Mental properties are thus *necessarily physically realised*. Second, whatever

particular physical property realises M is *necessarily sufficient* for M. This follows from the fact that (Supervenience) is a version of so-called *strong* supervenience.² Third, whatever particular physical property realises M might not have done so. (Supervenience) is an asymmetric necessitation relation. We can think of (Supervenience) as the shared commitment of all forms of physicalism, as a kind of *minimal physicalism*.³ Reductive physicalists, however, go further and identify mental properties with physical properties, whereas nonreductive physicalists insist that mental properties are irreducible⁴:

(Property Dualism) Mental properties are distinct from physical properties.

Now, one might think that since the problem about mental causation has traditionally been wedded to substance dualism, and nonreductive physicalism is incompatible with substance dualism, nonreductive physicalism should have no difficulties in handling this common sense phenomenon. But Kim⁵ has forcefully argued that all forms of dualism face this problem, regardless of how the mental is otherwise determined by the physical. This causal exclusion argument is *Descartes' revenge!* Moreover, Kim has argued that the only way for the nonreductive physicalist to respond to this argument is to reject completeness and accept emergentism including the possibility of downward causation.

This paper attacks both arguments. In Sec. III I show that Kim's other argument rests on a principle of property individuation that nonreductive physicalists should not accept, especially not those with functionalist leanings. This means that completeness is sound, hence that the exclusion principle upon which the causal exclusion argument depends must go. The justification for this principle derives from considerations about

overdetermination, but I argue in Sec. IV that given (Supervenience), there is a subtle way to avoid the undesirable consequences of overdetermination. The fact that this involves recognising a counterfactual dependence between the supposedly overdetermining causes has led some nonreductive physicalists to suggest that mental causation can be accounted for in terms of relevant counterfactuals. Sec. V contains some misgivings about this approach. The upshot is that although the nonreductive physicalist has the resources to resist both arguments, Kim's worries will linger on until she has provided some other account of mental causation. But let us first examine the causal exclusion argument. It proceeds in two stages corresponding to the fact that, disregarding physical-to-mental causation, mental causation splits up into mental-to-mental and mental-to-physical causation.

II. The Causal Exclusion Argument

Kim's first point is that mental-to-mental causation presupposes mental-to-physical causation. Consider:

(Mental Causation) Mental properties have causal powers.

What this means is that on occasion mental properties have the causal powers to bring about physical properties, as well as other mental properties.⁶ Think of a *causal power* as a function from circumstances *c* to causal effects *E*: the causal power of property *Q* consists in the effects *E* an object with *Q* would make in *c*.⁷ (Mental Causation) rejects *epiphenomenalism*, the view that the mental is physically caused, but can have no

physical or mental effects. It is highly implausible to think that mental properties are causally inert; or so I shall assume throughout (though see fn. 22). Voluntary action involves the causation of bodily movements by our beliefs and desires. It is a platitude that we tend to behave in a way that fulfils our desires if our beliefs are true.

Now, suppose mental property M causes mental property M*. (Supervenience) tells us that M* has physical base P*, so M and P* are both responsible for the occurrence of M*. P*, remember, is necessarily sufficient for M*. This means that M is jeopardised as a cause of M*. It would thus seem that the only way to save M as a cause of M* is by letting M directly cause P*. In fact, Kim thinks that all causation of supervenient properties must go via their base properties, or more precisely, that if you want to cause A, and A supervenes on B, *you must cause B*. A strong version of this principle says that if you want to cause A, and A supervenes on B, *you must knowingly cause B*. This means that you can bring about A only if you know which property B is, and you are able to act directly on B. Thus consider ethical and aesthetic supervenient properties: if you want to be a virtuous person, you must act on those of your natural properties that you know are necessarily sufficient for being virtuous, and if you want to make a beautiful sculpture, you must act on the constituting lump of marble knowing that a particular shape is necessarily sufficient for being beautiful. A weaker version of the principle says that if you want to cause A, and A supervenes on B, you must knowingly cause a property causally sufficient for B. This means that you can bring about B only if you know another property B* causally sufficient for B, and you are able to act directly on B*. Thus it is true that if I desire to relieve my headache, I have to act on the brain state upon which that mental state supervenes. There is no direct way – via telepathy,

telekinetics, or what have you - that I can ease my mental condition without intervening in my brain processes. That is why I take an aspirin. I have no knowledge of these processes, but even if I did, I would certainly not be able to act directly on them in the way I am able to act directly on a lump of marble. But I do know that in the right circumstances being an – ingested – aspirin is causally sufficient for relieving my discomfort. So my desire causes my behaviour in conjunction with my belief that aspirin normally has this effect. A better and weaker formulation of the principle that all causation of supervenient properties must go via their base properties, is that if you want to cause A, and A supervenes on B, *B must be caused*. This avoids potential confusion while ruling out spurious mental-to-mental causation. My desire to relieve my headache, and my belief that I can fulfil this desire by taking an aspirin, can only cause my arm to move, if they work through a physical causal chain starting from neuro-physiological correlates, going through the central nervous system, to a particular muscle contraction. Thus understood, it is credible that, given (Supervenience), the possibility of mental-to-mental causation hinges on the possibility of mental-to-physical causation.

What about mental-to-physical causation? Given that (Supervenience) tells us that M itself has physical base P, M and P both causally account for the occurrence of P*. But why think P is a cause of P* at all? Maybe this question has an *empirical* answer: neuro-scientists discover a continuous causal pathway connecting P with P*. But it also has a metaphysical answer to do with the following principle, which any *robust* physicalist should happily endorse:

(Completeness) Every physical effect has a sufficient physical cause.⁸

That is, any physical property that has a cause is completely fixed by purely physical causes. If the physical world is in this way *causally closed*, it must be in principle possible to give a complete physical account of all physical phenomena, where ‘in principle’ involves some sort of idealisation; it may well be that our cognitive abilities fall short, or that other available epistemic resources prove insufficient. If (Completeness) fails, such an account is in principle impossible since it will leave out some causes, like mental spirits, of certain physical phenomena.⁹

The problem is now that we have two different causally sufficient properties *competing* for the causation of P*, namely M and P:



On this background, Kim suggests that P pre-empts M as a cause of P*, so there is no causal work left for M to do. But what non-question-begging reason is there for thinking that P is *more* of a cause than M is when it comes to P*? We cannot a priori exclude the empirical possibility of a causal mechanism between M and P*. The reason derives from the following principle:

(Exclusion) If a property E has a sufficient cause C, then no other property C* distinct from C can be a cause of E.¹⁰

The idea is that if C is causally sufficient for E, then if C is instantiated, E must also be instantiated, whether or not any other property is instantiated. How then can a distinct property C* also be a cause of E? To allow for both C and C* to be causally sufficient for

E, is to allow for causal *overdetermination*, and it is hard to believe that *all* cases of mental-to-physical causation are cases of overdetermination. There are at least three reasons for this. First, causation typically involves counterfactual dependence. But if P* is overdetermined by M and P, it would, for instance, not be true that if M had not occurred, P* would not have occurred either; P* would have occurred anyway, thanks to P. Second, causes typically raise the chances of their effects. But if P* is overdetermined by M and P, $ch_M(P^*) = ch_{\sim M}(P^*)$, and $ch_P(P^*) = ch_{\sim P}(P^*)$. This makes M a bad candidate for a cause of P*, and similarly with P and P*. Yet bad candidates sometimes win. The point is not that causation entails counterfactual dependence and chance-raising, but that intuitively there is a *prima facie* reason to believe that a case without the latter is a case without the former. Third, if P* has two independent, sufficient causes M and P, P* also has two independent and complete, causal explanations. But if, for instance, P is capable of fully accounting for the occurrence of P*, no explanatory work is left for the independent M to do.¹¹

Given (Exclusion), it is now easy to see why P excludes M as a cause of P*. (Completeness) says that P* has sufficient physical cause P, and (Exclusion) then says that P* can have no other cause distinct from P, and so, in particular, P* cannot have M as a distinct cause. But if P excludes M as a cause of P*, M cannot be a cause of M*, because, as we argued, M can only cause M* if M can cause P*. Given that we have made no special assumptions about which mental property M is, the upshot is that nonreductive physicalism cannot accommodate (Mental Causation).

III. Downward Causation

It is important to get clear on the dialectic. What the causal exclusion argument shows is, strictly speaking, that the conjunction of (Supervenience), (Property Dualism), (Mental Causation), (Completeness), and (Exclusion) is inconsistent, hence that at least one must go. In particular, denial of (Supervenience) is little help since the remaining four are still inconsistent. This shows that substance dualism in this respect is no worse off than nonreductive physicalism. The problem for the latter is that all five tenets seem to be warranted. (Property Dualism) and (Supervenience) define nonreductive physicalism¹², (Mental Causation) is, we take it, indispensable on any viable account of human agency, (Completeness) is supposedly common ground between all forms of robust physicalism, and (Exclusion) is a principle we all seem to accept as an integral part of the causal structure of reality.

Suppose the nonreductive physicalist chooses to blame the argument on (Completeness). Then we get the following picture:



M and M* supervene on P and P* respectively, and M causes M* by causing M*'s base property P*, which is necessarily sufficient for M*. We thus still assume that mental-to-mental causation must go through mental-to-physical causation. Giving up on (Completeness), but holding on to (Exclusion), need not entail the possibility of

irreducible mental-to-mental causation. But it does mean that if M is causally sufficient for P*, P cannot also be causally sufficient for P*. This raises the following question: is this position a form of physicalism? Given that we have defined ‘minimal physicalism’ as (Supervenience), the answer would seem to be affirmative. But we have also assumed that robust physicalists are reluctant to give up (Completeness). In fact some of them regard (Completeness) as being scientifically verified by discoveries made in physics in the course of the twentieth century.¹³ In any case, the interesting issue is not about terminology, but plausibility. Is the combination of (Property Dualism), (Supervenience), (Mental Causation), (Exclusion) and not-(Completeness) an attractive picture? Well, on this view, M is best seen as an *emergent property*: a genuinely novel kind of property of a whole consisting of parts of an old kind that emerges, not because something from the outside is added, but when and only when those parts are put together in the right kind of way. Crucially, the causal power of an emergent property is distinct from, or irreducible to, the causal powers of the lower-level properties on which it, in some sense, depends.¹⁴ This is precisely what this combination allows for. Mental properties are distinct from, but supervene on, physical properties, and they are causally sufficient for physical effects, which themselves lack sufficient physical causes. If (Completeness) were true, these effects would have sufficient physical causes, which would then, by (Exclusion), pre-empt the causal powers of the mental properties. Thus mental properties have causal powers over and above the causal powers of their physical base properties.

According to *emergentism*, our world is a *layered world*: there is a hierarchy of distinct yet connected levels starting from the physical level. Specific to each level, there are distinct kinds of substances wholly composed of kinds from lower-levels all the way

down to elementary material particles. Each kind has specific properties in virtue of a characteristic organizational complexity, and some of these properties will have emergent causal powers. What is more, there are special *emergent laws*, neither reducible to, nor derivable from, lower-level laws, which attribute these causal powers to the types of properties in question.¹⁵

The difficulty, or challenge if you like, that arises in connection with emergentism is how higher-level properties can both cause, and be supervenient on, lower-level properties. How can it be that:

“mentality, having emerged from physical-biological processes, takes on a causal life of its own and begins to exercise causal influence “downward” to affect what goes on in the underlying physical-biological processes”?¹⁶

In short, how can *upward determination* from the physical to the mental be combined with *downward causation* from the mental to the physical? Old-style English emergentists were explicitly committed to the idea of downward causation:

“... when some new kind of relatedness is supervenient (say at the level of life), the way in which the physical events which are evolved run their course is different in virtue of its presence – different from what it would have been if life had been absent.”¹⁷

On the face of it, there is nothing spurious about downward causation *as such*. It is not *prima facie* incoherent to imagine that M can be causally sufficient for P* while at the same time supervene on P¹⁸. The trouble begins only when P* is not sufficiently caused by physical property P. My headache causes a desire for pain-relief. Both supervene on distinct neuro-physiological states of my brain. How can the former act directly on the

base property of the latter without there being some other neuro-physiological causal influences? It would seem equally if not more natural that the base property of my desire is sufficiently caused by the base property of my pain.¹⁹

Are there examples of downward causation from the natural sciences? Take Ohm's Law that the voltage V in a circuit is a function of the current I flowing in the circuit and the resistance R to this flow: $V = I \times R$. Now suppose we double the total resistance in a circuit with constant voltage. The amount of current flowing in such a circuit is inversely proportional to its total resistance, so this means that we halve the current of the circuit. Here a macroscopic event – inserting another light bulb – presumably causes a microscopic event – change in movement of electrons through the conductor, while also causing another macroscopic event - causing the current to be half of what it was before the insertion. Obviously, the macroscopic facts about the strength of the current supervene on the microscopic facts about the motion of electrons, but the emergentist can allow for that.

I do not want to take a stand on whether we can model mind-body downward causation on examples like this. What I do want to argue is that Kim is wrong in thinking that nonreductive physicalism *entails* the possibility of downward causation. Note first that nonreductive physicalism is a form of *mental realism*: mental properties are real, and not just heuristic devices for making convenient predictions. But one should arguably not postulate the existence of concrete entities unless they actually or possibly make a causal difference in the physical world; in particular, no instantiated concrete properties should be causally inert. Consider:

(Alexander's Dictum) If A is a real property, then A has causal power.

The reason for the name is that Alexander [1920, p. 8] famously wrote the following in a striking critique of epiphenomenalism:

“[Epiphenomenalism] supposes something to exist in nature which has nothing to do, no purpose to serve, a species of noblesse which depends on the work of its inferiors, but is kept for show and might as well, and undoubtedly would in time be abolished.”

The line of thought is that we have no reason to believe in the existence of mental properties if, as epiphenomenalism states, they are entirely causally impotent. If you removed all mentality from an epiphenomenalist world, you would not change a single physical fact about it. Thus epiphenomenalism leads ultimately to *mental irrealism*. But contrast with the following principle about the individuation of properties by their causal powers²⁰:

(Causal Individuation) If A is an irreducible property, then A has irreducible causal power.

Restrict A to concrete properties instantiated in the actual world. Now suppose A is supervenient, and has B as its only distinct base property; if A has multiple base properties, take B to be a disjunction of these. What the principle says is then that A can make a causal difference, B cannot actually or possibly make, for if the causal difference that A can make is one that B can actually or possibly make, then A's causal power is reducible to the causal power of B. A therefore has a causal power that is irreducibly distinct from the causal power of B, hence A is emergent. In other words, (Causal

Individuation) forces the nonreductive physicalist to abandon (Completeness), and embrace emergentism with the possibility of downward causation. Here is Kim²¹:

“...the emergentist and nonreductive physicalist are mental realists, and Mental Realism, via Alexander’s dictum, entails causal powers for mental properties... mental properties, on both positions are irreducible net additions to the world. And this must mean, on Alexander’s dictum, that mental properties bring with them *new causal powers, powers that no underlying physical-biological properties can deliver*. For unless mentality made causal contributions that are genuinely novel, the claim that it is a distinct and irreducible phenomena would be hollow and empty. To be real, Alexander has said, is to have causal powers; *to be real, new, and irreducible, therefore, must be to have new, irreducible causal powers.*”

We can regiment this line of thought as follows:

- | | |
|--|----------------------|
| (1) Mental properties are real | (Mental realism) |
| (2) So, mental properties have causal powers | (Alexander’s Dictum) |
| (3) But mental properties are irreducible properties | (Property Dualism) |
| (4) So, mental properties have irreducible causal powers | (Alexander’s Dictum) |

The problem is that (Alexander’s Dictum) does not license the step from (3) to (4); only (Causal Individuation) would validate this move. All the dictum says is that to be a real property is to have causal power, regardless of whether that property or its causal power is reducible or not. It is one thing to say that to be a real property is to confer causal power on its instances, another thing to say that to be a real property is to confer *irreducible* causal power on its instances. The nonreductive physicalist will accept the former, but need not agree to the latter.²²

For example, some nonreductive physicalists maintain that the reason mental properties are distinct from, yet dependent on, physical properties, is that they are multiple realisable. On their view, mental properties are second-order properties defined over first-order physical properties. To have mental property M is to have the property of having a property P that typically plays causal role R. M is not identical to P since distinct Ps play R in different systems. Nor is M identical to a disjunction of Ps. M is disjunctively realised, not a disjunctive property.²³ If P plays R in a particular system s, having P is one way of having M. When this instance of P realises M in s, having P causing P* is therefore having M causing P*. M in s is thus a cause of P* in virtue of one of its realiser properties P causing P*. On this view, M and P are distinct, yet they do not compete for the causation of P*. Thus if M is realised by an instance of P in s, the causal power of this instance of P is identical to the causal power of M in s; the causal power of M is a proper subset of the causal power of P *simpliciter*, if instances of P can realise higher-order properties other than M. So, M is a causally efficacious property, but not a property with its own irreducible causal power. In fact, Kim elsewhere (1998, p. 54) defends the claim that an instance of a higher-order mental property inherit all its causal power from the first-order physical property that realises it. (Causal Individuation), however, requires precisely that the causal power of the former exceed the causal power of the latter.

Bearing this realisation view in mind, it is not “hollow and empty”²⁴ to claim that mentality is a distinct and irreducible phenomenon unless it makes an irreducible causal contribution to the world. It surely would be “hollow and empty” to make such a claim if one were not ready to accept (Alexander’s Dictum), but the nonreductive physicalist is

not an epiphenomenalist. Kim seems to believe that (Causal Individuation) somehow follows from considerations about the implausibility of epiphenomenalism if mental properties are irreducible, but it calls for independent support. Only property dualists who are already committed to the possibility of downward causation are inclined to accept (Causal Individuation). In short, it is Kim's failure to distinguish (Alexander's Dictum) from (Causal Individuation) that leads him to think that downward causation *follows from* nonreductive physicalism.²⁵

IV. Causal Exclusion

We have defined *minimal physicalism* as (Supervenience), but also maintained that any *robust* form of physicalism should endorse (Completeness). It follows that nonreductive, robust physicalism must deny (Exclusion) if (Mental Causation) is taken on board. The problem raised by the causal exclusion argument is, in Kim's [1998, p. 30] words: "Given that every physical event that has a cause has a physical cause, how is a mental cause also possible?" A mental cause is only possible if either it is identical to the physical cause, or it is not excluded by the physical cause. Given (Property Dualism), the robust nonreductive physicalist must maintain that not every caused physical property has only physical causes. But if the physical effect P* has both a physical cause P and a distinct mental cause M, P* is *causally overdetermined*. In short, (Property Dualism), (Mental Causation) and (Completeness) entail overdetermination. Yet we agreed that not all cases of mental-to-physical causation can be cases of overdetermination.

Suppose two expert assassins both actually kill a convict by simultaneously shooting him through the head. I gave three reasons for suspicion about such putative cases of overdetermination. First, if the one assassin had shot but not killed the convict, or even if he had failed to shoot, the other assassin would have shot and killed him (at the same time and in the same way). Second, neither of the assassins' shots, taken individually, increases the chance of the convict's death. Third, if the shots are independent and each sufficient for the death of the convict, we have two complete and independent causal explanations of his death. Let us call this:

(Overdetermination): E is overdetermined by C and C* iff (i) C is sufficient for E, (ii) C* is sufficient for E, (iii) if C occurred without C*, E would have occurred, and (iv) if C* occurred without C, E would have occurred.

Thus in our example each shot is sufficient for the convict's death, and each shot would occur and cause the convict's death without the other. (i)-(iv) are therefore all *non-vacuously true*.²⁶ But it is now clear that P* is not overdetermined by M and P in this way. We have defined (Supervenience) in the strong sense such that if P is M's base property, P is necessarily sufficient for M. This means that P is counterfactually dependent on M: $\sim M \Box \rightarrow \sim P$. There are no close worlds in which M fails to occur but P does, i.e. all close not-M worlds are not-P worlds.²⁷ So, M and P are each sufficient for P*, but P would occur only if M did also occur. This means that the counterfactual $(P \ \& \ \sim M) \Box \rightarrow P^*$ is not non-vacuously true as it would be if P* were overdetermined by P and M. The point about vacuity is important. In my view, what the nonreductive physicalist should claim is not that $(P \ \& \ \sim M) \Box \rightarrow P^*$ is false. First, in order to allow the antecedent to be true, she would have to weaken the supervenience relation. Second, the question

arises: if P^* had not occurred, had P occurred without M , how could P be sufficient for P^* ? P would seem to be sufficient for P^* only if M did also occur. In general, denial of (iii) seems to be in tension with (i), and likewise with (iv) and (ii). Third, on the assumption that M is irreducibly mental, the denial of $(P \ \& \ \sim M) \ \Box \rightarrow P^*$ may jeopardise (Completeness); P 's sufficiency for P^* , remember, is warranted by this principle.

Fortunately, what matters in the present context is that mental causation by the nonreductive physicalist's lights is not like the standard examples of overdetermination such as the execution of the convict. In the latter case, (iii) and (iv) are both non-vacuously true, but in the former case one of them is vacuous. So, although I argue in Sec. V that the nonreductive physicalist should accept the non-vacuous truth of the *other* counterfactual $(M \ \& \ \sim P) \ \Box \rightarrow P^*$, she still has the resources to resist the undesirable consequences of overdetermination. To be clear, the nonreductive physicalist thus strictly speaking accepts the truth of (i)-(iv), and so is committed to (Overdetermination). But this commitment is innocuous if either (iii) or (iv) is vacuously true. The nonreductive physicalist need therefore only endorse an acceptable form of overdetermination.²⁸

What about the counterfactual dependency of M on P ? Given that M is multiple realisable, it is possible that M occurs without the *particular* base property P , so $\sim P \ \Box \rightarrow \sim M$ is false. But if $\cup P$ is the disjunctive property of having either M 's actual base property P , or one of its possible base properties, $\sim \cup P \ \Box \rightarrow \sim M$ is true. Kim thinks differently. He [1998, p. 45] asks us to:

“...consider a world in which the physical cause does not occur and which in other respects is as much like our world as possible. The overdetermination approach says that

in such a world, the mental cause causes a physical event – namely, that the principle of causal closure of the physical no longer holds.”

On the present account, $\diamond\exists x(Mx \ \& \ \sim Px)$. But this does not mean that in the absence of P, M would be the sole cause of P* in breach of (Completeness)? The closest not-P worlds are presumably worlds in which M does not occur at all; or alternatively worlds in which M has an alternative physical base. (Supervenience) says that it is *metaphysically* impossible that something has M without a physical base property, $\forall x(Mx \rightarrow \cup Px)$. Kim’s objection would be good only if P* is overdetermined by P and M in the same way that the convict’s death is overdetermined by the assassins’ shots for only then would M cause P* in the absence of a physical base property. So, again Kim thinks that nonreductive physicalism entails the possibility of downward causation, but this time by making a false assumption about causation.²⁹

The nonreductive physicalist can thus justifiably reject (Exclusion) when M and P are counterfactually dependent.³⁰ First, it is not the case, as we have just seen, that if M does not occur, P will single-handedly cause P*. If P occurs, M will necessarily also occur. Second, the occurrence of M does raise the chance of the occurrence of P*, that is $ch_M(P^*) > ch_{\sim M}(P^*)$. It is only if P caused P* independently of M that $ch_M(P^*) = ch_{\sim M}(P^*)$. Third, if M and P are counterfactually dependent, they do not provide two complete and independent causal explanations of why P* occurred. The explanation of how M caused P* will mention the fact that M has a physical base property P, which is also a cause of P*, and the explanation of how P caused P* will mention the fact that P is necessarily sufficient for M, which is also a cause of P*.³¹

There are, as Block [2003] argues, independent reasons for scepticism about (Exclusion). Take a typical case of a higher-level property causing another higher-level property, e.g. being a storm causing being a flood on the coast. Now, there is a lower-level, highly complex property to do with the motion and distribution of air molecules that is an equally good candidate for a cause of the flood. By (Exclusion) we can show that the latter pre-empts the former. But the latter is itself a higher-level property with respect to an at least equally complex property at the atomic level, and then we can use (Exclusion) once more to show that this new property pre-empts the molecule property, and so on. Note that (Completeness) plays no role in this argument since all causes and effects are physical causes and effects. This confronts us with the following dilemma: if there is a bottom level of elementary particles in physics, all other causation drains down to this level; if there is not, all causation drains down to a bottomless nothing!

Kim's response [2003, pp. 168, 175] is that identification of the competing causes at an appropriate level stops the drainage. If the identification concerns properties themselves, then the exclusion problem returns on the assumption that the higher-level properties in question are multiple realisable and that the relevant identifications therefore are taken to involve disjunctive properties. Suppose special science property S causes E . Then if $S = \cup P$, one of the disjuncts P' will presumably be causally sufficient for E . But $\cup P$ and P' are distinct as they are not necessarily properties of all and only the same objects. So, by (Exclusion), we can show that P' pre-empts S 's causal power. Note that whether the level at which S occurs is causally closed plays no role in this argument. If, on the other hand, only instances of higher- and lower-level properties are identified, there is no question about the causal power of the former draining down to the latter, but

in that case the causal exclusion argument provides no reason to believe in property identities; in fact this is in line with the realisation view as sketched in Sec. IV.

V. Counterfactual Dependence

The foregoing brings up the question: in what sense are M and P causes of P*? To give a detailed, positive account of mental causation is beyond the scope of this paper. One desideratum is that M and P are *causally homogeneous*.³² Thus Loewer [2001a; 2001b] and Crane [2001a, 2001b] have proposed that the nonreductive physicalist can facilitate (Mental Causation) if a Lewis-style counterfactual analysis of causation is adopted. The rest of the paper contains some worries about this proposal. The thought is roughly that C causes E if $\sim C \Box \rightarrow \sim E$ and $C \Box \rightarrow E$. If we suppose that C and E are *actually instantiated* properties, the latter is trivially true. And in order to exclude metaphysical, conceptual and other non-causal yet counterfactual dependencies, they require that C and E be *genuinely distinct* properties. So, on this picture, P counts as a cause of P*: $\sim P \Box \rightarrow \sim P^*$. But M also counts as a cause of P* even though there is no downward causation: both $\sim M \Box \rightarrow \sim P$ and $\sim P \Box \rightarrow \sim P^*$ are true, so by transitivity, $\sim M \Box \rightarrow \sim P^*$ is true.³³ M is a cause of P*, because there is a chain of counterfactual dependencies from M to P*. So, P and M are causes in the *same sense* that if either had not occurred, P* would not have occurred.³⁴

As is familiar, there are a number of objections to this account, some of which are specifically aimed at mental causation while others are more general. Set the latter aside,³⁵ and focus on the potentially most worrying objection, namely that this account

fails to rule out the epiphenomenalist alternative according to which the physical causes the mental, but the mental causes nothing. Kim's charge [1998, p. 71] is that the present account cannot distinguish epiphenomenal from genuinely causal counterfactuals. How do we know that mental properties are not like the shadows cast by a moving car, or the symptoms of a disease in progress. Thus suppose P and P* are the locations of a car at time t_1 and t_2 respectively, and that M and M* are the shadows thrown by the car at t_1 and t_2 respectively. If that were so, the counterfactuals $\sim M \Box \rightarrow \sim M^*$ and $\sim M \Box \rightarrow \sim P^*$ would still be true, but M would neither be a cause of M* nor of P*. The challenge is to explain in terms of relevant counterfactuals why mental properties are not epiphenomenal danglers from their physical groundings!

According to epiphenomenalism, there are law-like correlations between M and P. If mental causation were modelled on the car/shadow and disease/symptom examples, this would mean that M was caused by P, or alternatively, that M supervenes on P with nomological necessity. In any case, what is distinctive about epiphenomenalism is that P pre-empts M with respect to P*. One way of defending the counterfactual analysis is thus to add that M causes P* only if M is not pre-empted by its base property P with respect to P*. Loewer [2001a; 2001b, pp. 322-3] suggests that a sufficient condition for P to pre-empt M with respect to P* is: $\sim P \Box \rightarrow \sim P^*$ and $(P \ \& \ \sim M) \Box \rightarrow P^*$. Both conditions are met on the epiphenomenalist account. We all agree that $\sim P \Box \rightarrow \sim P^*$ is true. And the epiphenomenalist also allows not only that P & $\sim M$ is metaphysically possible, but also that the change in laws of nature/particular physical facts it takes to make P & $\sim M$ & P* true in a world is smaller than the change in laws of nature/particular physical facts it takes to make P & $\sim M$ & $\sim P^*$ true in a world. The closest worlds in which the car occurs

at t_1 without the shadow at t_1 are also worlds in which the car occurs at t_2 . But, as we saw, the nonreductive physicalist denies that $P \ \& \ \sim M$ is metaphysically possible, so, she can at least dispute the epiphenomenalist's claim that $(P \ \& \ \sim M) \ \Box \rightarrow P^*$ is non-vacuously true, and this may be enough to show that P does not pre-empt M with respect to P^* .

The problem with this response is that the sufficiency claim for pre-emption can be used to show that M pre-empts P with respect to P^* , so that M and P are not after all causally homogeneous. The first condition $\sim M \ \Box \rightarrow \sim P^*$ is met by transitivity from $\sim M \ \Box \rightarrow \sim P$ and $\sim P \ \Box \rightarrow \sim P^*$, and the second condition $(M \ \& \ \sim P) \ \Box \rightarrow P^*$ is arguably also non-vacuously true given that the closest $M \ \& \ \sim P$ worlds are worlds in which M has an alternative base property. Pace Kim, neither (Completeness) nor (Supervenience) needs to be violated in order to find an $M \ \& \ \sim P \ \& \ P^*$ world. So, the question is: are the worlds in which M and its alternative base property cause P^* closer to the actual world than the worlds in which they do not? I think so. Clearly, some $M \ \& \ \sim P \ \& \ P^*$ worlds are more distant than some $M \ \& \ \sim P \ \& \ \sim P^*$ worlds, namely those in which M has a very different physical base property. But that is irrelevant. We should only look at those worlds in which M has a *different, but very similar*, physical base property, and in those worlds that property will still cause P^* . If pain had supervened on a physical property, different from, but very similar to, C-fibre stimulation, pain and that property would still cause a desire for pain relief by causing the base property of that desire.³⁶

But even if we are able in some non-*ad hoc* way to keep out the epiphenomenal counterfactuals, a worry remains about the allegedly causal counterfactuals themselves. The thought was that $\sim M \ \Box \rightarrow \sim P^*$ is true by transitivity from $\sim M \ \Box \rightarrow \sim P$ and $\sim P \ \Box \rightarrow \sim P^*$. But we would only have reason to believe that the former is causal if both of the

latter are also causal. The point is general. One can infer $A \Box \rightarrow C$ of type x from $A \Box \rightarrow B$ and $B \Box \rightarrow C$ only if they are both also of type x . Compare with: if there were watery stuff in the glass, there would be water in the glass; if there were water in the glass, there would be H_2O in the glass; so, if there were watery stuff in the glass, there would be H_2O in the glass. The reasoning is not truth-preserving: use of transitivity of counterfactuals should not conflate conceptual with metaphysical relations.³⁷ Similarly, $\sim M \Box \rightarrow \sim P^*$ is causal only if $\sim M \Box \rightarrow \sim P$ and $\sim P \Box \rightarrow \sim P^*$ are also both causal. But the relationship between M and P is not happily thought of in causal terms. Supervenience is not a kind of causation. Causal relations (normally) have intermediate causal stages, but P and M do not, and causes (actually) precede their effects, but P and M are instantiated simultaneously. Moreover, it is implausible that P causes M – P causes P^* and is necessarily sufficient for M – but it is even more implausible that M causes P .

Let us dwell on Lewis' modified account [2000] for a moment. Lewis defined causation as a chain consisting of a finite sequence of actual events where each event causally depends on an antecedent event in the sequence, and where causal dependence is counterfactual (E causally depends on C if $\sim C \Box \rightarrow \sim E$).³⁸ In short, a chain of counterfactual dependences is sufficient for causation. Both causation and causal dependence are transitive, so M can be a cause of P^* if there is a causal chain leading from M to P^* . But we have seen no reason to believe in the existence of such a chain since P is not intuitively caused by M . So, even if we assumed that a suitably modified, counterfactual theory of causation were able to deal with all the general difficulties, there would still remain specific problems about pre-emption and the causal nature of the relevant counterfactuals in the case of mental-to-physical causation.

Let us take stock. One way to read the causal exclusion argument is that if we hold on to (Property Dualism) as our nonreductive physicalist does, then we can save (Mental Causation), only if either (Completeness) or (Exclusion) is given up. We have argued in Sec. III that there is no reason why we should deny (Completeness) and embrace the possibility of downward causation. Kim's additional argument for this claim makes assumptions about the causal power of irreducible mental properties, which no nonreductive physicalist need accept. Instead we argued in Sec. IV that the nonreductive physicalist should reject (Exclusion), which is independently implausible, and hence accept (Overdetermination). But given the way (Supervenience) is cashed out, the counterintuitive consequences of some cases of an effect having two sufficient causes can be avoided by insisting that they be counterfactually dependent. To suggest the best response to the causal exclusion argument on behalf of nonreductive physicalism is however not to say anything positive about what mental causation might be on this view. The counterfactual analysis meets the homogeneity constraint on mental and physical causation, but is beset with severe difficulties. For instance, it owes an account of why mental causes are not epiphenomenal.³⁹ I believe the nonreductive physicalist must deliver some account of mental causation in terms other than counterfactuals that not only respects this constraint, but also entails sufficient dependency between distinct mental and physical causes to sustain a respectable form of overdetermination.⁴⁰ Until the nonreductive physicalist has satisfactorily advanced such an account, she may be able to respond to the causal exclusion argument in the way I have recommended, but Kim's worries about the causal efficacy of the mental will hang on.⁴¹

References

- Alexander, S. (1920), *Space, Time and Deity*, London: Macmillan.
- Beckerman, A., Flohr, H., and Kim, J. (eds.) (1992), *Emergentism and Reduction*. New York and Berlin: De Gruyter.
- Bennett, K. (2003), 'Why the Exclusion Problem Seems Intractable, and How, Just Maybe, To Tract It', *Nous* 73:3, pp. 471-497.
- Block, N. (2003): 'Do Causal Powers Drain Away?', *Philosophy and Phenomenological Research*, Vol. LXVII, No. 1.
- Chalmers, D. (1996), *The Conscious Mind: In Search of a Fundamental Theory*, Oxford: Oxford University Press.
- Crane, T. (1995), 'The Mental Causation Debate', *Proceedings of Aristotelian Society*, Supp. Vol.
- (2001a), *Elements of Mind. An Introduction to the Philosophy of Mind*, Oxford: Oxford University Press.
- (2001b), 'The Significance of Emergence', in Gillett, C. and Loewer, B., (eds.), *Physicalism and Its Discontents*.
- Davidson, D. (1980), 'Mental Events', in his *Essays on Actions and Events*, Oxford: Clarendon Press.
- (1993), 'Thinking Causes', in Heil, J. and Mele, A., eds.: *Mental Causation*.
- Fodor, J. (1997), 'Special Sciences; Still Autonomous After All These Years', in Tomberlin, J. (ed.), *Philosophical Perspectives*, v. 11, Oxford: Blackwell Publishers.

Gillett, C. and Loewer, B. (eds.) (2001), *Physicalism and Its Discontents*, Cambridge: Cambridge University Press.

Heil, J. and Mele, A. (eds.) (1993), *Mental Causation*, Oxford: Clarendon Press.

Horgan, T. (1984), 'Supervenience and Cosmic Hermeneutics', *Southern Journal of Philosophy*, 22, suppl., pp. 19-38.

————— (1993), 'From Supervenience to Superdupervenience: Meeting the Demands of a Material World', *Mind*, 102:555-86.

————— (1997), 'Kim on Mental Causation and Causal Exclusion', *Philosophical Perspectives* 11, pp. 165-84.

Jackson, F. and Pettit, P. (1990), 'Program Explanation: a general perspective', *Analysis* 50, 107-117.

Kim, J. (1992), 'Downward Causation', in Beckerman, A., Flohr, H., and Kim, J. (eds.), *Emergentism and Reduction*.

————— (1993), *Supervenience and Mind: Selected Essays*, Cambridge: Cambridge University Press.

————— (1998), *Mind in a Physical World*, Cambridge: Cambridge University Press.

————— (1999), 'Making Sense of Emergence', *Philosophical Studies*, 95.

————— (2001), 'Mental Causation and Consciousness: The Two Mind-Body Problems for the Physicalist', in Gillett, C. and Loewer, B. (eds.), *Physicalism and Its Discontents*.

————— (2003), 'Blocking Causal Drainage and Other Maintenance Chores with Mental Causation', *Philosophy and Phenomenological Research*, Vol. LXVII, No. 1.

Lewis, D. (1973a), *Counterfactuals*, Oxford: Blackwell Publishers.

————— (1973b), 'Causation', *Journal of Philosophy*, 70.

- (2000), 'Causation as Influence', *Journal of Philosophy*, 97.
- Loewer, B. (2001a), 'From Physics to Physicalism', in Gillett, C. and Loewer, B. (eds.), *Physicalism and Its Discontents*.
- (2001b), 'Book Review: *Mind in a Physical World*', *The Journal of Philosophy*, pp. 315-24.
- McLaughlin, B. (1992), 'The Rise and Fall of British Emergentism', in Beckerman, A., Flohr, H., and Kim, J., eds.: *Emergentism and Reduction*.
- (1995), 'Varieties of Supervenience', in Savellos, E. and Yalcin, U., eds.: *Supervenience: New Essays*.
- Mellor, D.H. (1995), *The Facts of Causation*, London: Routledge.
- Menzies, P. (1996), 'Probabilistic Causation and the Pre-emption Problem', *Mind* 105, pp. 85-117.
- (2001), 'The Causal Efficacy of Mental States', in *The Structure of the World: the renewal of Metaphysics in the Australian School*, (ed.) J.M. Monnoyer.
- Monnoyer, J. M. (ed.) (2001), *The Structure of the World: the renewal of Metaphysics in the Australian School*, Vrin Publishers.
- Morgan, C.L. (1930), *Emergent Revolution*, London: Williams and Norgate.
- Papineau, D. (2001), 'The Rise of Physicalism', in Gillett, C. and Loewer, B., eds.: *Physicalism and Its Discontents*.
- (2002), *Thinking about Consciousness*, Oxford: Oxford University Press.
- Pereboom, D. (2002), 'Robust Nonreductive Materialism', *Journal of Philosophy*, 99.
- Shoemaker, S. (1980), 'Causality and Properties', in Van Inwagen, P. ed.: *Time and Cause*.

Savellos, E. and Yalcin, U. (eds.) (1995), *Supervenience: New Essays*, Cambridge: Cambridge University Press.

Van Inwagen, P. (ed.) (1980), *Time and Cause*, Dordrecht: D. Reidel Publishing.

Yablo, S. (1992), 'Mental Causation', *The Philosophical Review*, Vol. 101, No. 2.

¹ Opponents of physicalism will typically not deny that mental properties supervene on physical properties with nomological or natural necessity; see Chalmers [1996].

² This is indicated by the second occurrence of 'necessarily': if something has M in virtue of having P, then anything else that has P has M even if not in the actual world. Formally:

$$[\exists x (Mx) \rightarrow \exists P(Px) \ \& \ \forall y (Py \rightarrow My)].$$

³ Kim [1998, pp. 14-5]. Physicalism is typically formulated as a global supervenience thesis, involving quantification over worlds rather than objects and properties. The chief reason for this has to do with wide content mental states. But (Supervenience) can facilitate such states if only their supervenient bases include relational or extrinsic physical properties. For details, see Kim [1993, pp. 85-9; 161-65] and McLaughlin [1995].

⁴ This is not to say that the nonreductive physicalists have nothing to offer as to why mental properties supervene on physical properties. Horgan [1984; 1993] and Kim [1993; 1998] have repeatedly pointed out that the nonreductive physicalist must cite some

metaphysical notion – ‘superdupervenience’ – that grounds (Supervenience). One such notion on offer is realisation; Kim [1998, pp. 23-24], and fn. 39.

⁵ See his [1993, pp. 350-7; 1998, pp. 39-47; 2001, pp. 275-8].

⁶ I shall follow Kim [1998, p. 41] in talking conveniently about a property F causing property G, but this can be taken as short for talk of an instance of F causing an instance of G in virtue of the fact that the first is an F-instance and the second is a G-instance. If one follows Davidson [1970, 1993] in thinking that causes are events, understood not as property exemplifications at times, but as basic particulars abstracted from any properties they may have, one will have to accept that no event is causally efficacious in virtue of its mental properties. I agree with Kim [1993, pp. 268-71] and others that it is implausible that the mental qua mental should make no causal difference in the physical world.

⁷ Shoemaker [1980, p. 255].

⁸ By ‘causally sufficient’ I mean throughout ‘causally sufficient given the laws of nature and other causally relevant circumstances’. Note also that if determinism fails, we must reformulate the principle: for every physical effect, there is a physical event, or series of physical events, which determine the objective probability of that effect. Nothing in the argument hinges on this issue.

⁹ Kim [1998, p. 40] expresses the causal closure of the physical domain in this way: “If you pick any physical event and out its causal ancestry and posterity, that will never take you outside the physical domain. That is, no causal chain will ever cross the boundary between the physical and the non-physical.” It follows that (Completeness) is compatible with the existence of non-physical properties outside the physical domain, and even with the obtaining of causal relations between them. Nor does (Completeness) rule out the possibility of physical properties lacking sufficient physical causes; all it says is that if such properties are caused, they have a sufficient physical cause.

¹⁰ Kim [2001, p. 276]. For clarity I leave out time indices. Reflect that (Exclusion) could not be used to show that P* excludes M as a cause of M*, because P* is not happily thought of as a cause of M*. To show this, one would need a more general principle to the effect that if E is determined by C, then E is not determined by C*, distinct from C; see Kim [2001, pp. 276-7; 1993, pp. 237-64]. The worry about this principle is that it

excludes the possibility of M having a physical cause distinct from its physical base property, e.g. tissue damage causes pain, which supervenes on C-fibre stimulation.

¹¹ Kim [1998, p. 45; 1993, pp. 239, 247, 250-3] formulates a principle for explanatory exclusion: no event can be given more than one complete and independent causal explanation. It is not less implausible that M and P are each individually necessary, but only jointly sufficient for P*; think of the spark, gas and oxygen in a car engine cylinder together causing an explosion. (Mental Causation) and (Completeness) say that M and P, respectively, are individually sufficient for P*.

¹² The reductive physicalist rejects (Property Dualism), and so can consistently endorse the remaining four tenets - on this view (Supervenience) is trivially true. If mental and physical properties are identical, causation at the mental level is just causation at the physical level. Kim [2003, p. 165] now believes that (Property Dualism) is the premise assumed for *reductio*. However, see Sec. IV for some worries about the claim that the argument provides a strong case for such property identities.

¹³ Papineau [2001, pp. 3-37; 2002, pp. 44-46, Appendix I].

¹⁴ It is also often assumed that emergent properties are theoretically unpredictable and inexplicable from knowledge of those parts and how they are put together. Emergent properties contrast with resultant properties which are predictable, e.g. having a certain mass. Crane [2001b, pp. 211-16] and Pereboom [2002, pp. 509-10] argue that these ideas about emergence amount to nothing the nonreductive physicalist should not accept.

¹⁵ For more on the history of emergentism see McLaughlin [1992].

¹⁶ Kim [1993, p. 349]. See also his [1998, pp. 54-5].

¹⁷ Morgan [1923, p. 16].

¹⁸ McLaughlin [1992, pp. 53-4, 74-5] argues that downward causation is compatible with the laws of mechanics, quantum mechanics and relativity theory.

¹⁹ There are, as Kim points out, different possible kinds of downward causation. Reflexive downward causation says that M causes its own base property P, and synchronic reflexive downward causation says that M causes P at the same time at which P brings about M. Although Kim [1999, p. 26] finds that “there need not be anything strange or incoherent in the idea of downward causation as such”, the latter involves a

metaphysically odd, causal loop. But, as fig. 2 highlights, the emergentist is only committed to diachronic, irreflexive downward causation.

²⁰ See also Shoemaker [1980] who defines a property as the potential for contributing to the causal powers of objects that have it. Shoemaker's theory has the - in my view implausible - consequence that all of the causal potentialities possessed by a property are essential to it, that is belong to it in all metaphysically possible worlds.

²¹ [1993, p. 350]. See also his [1999, p. 5].

²² At least when A is an intentional property. Some nonreductive physicalists take a hybrid view according to which phenomenal properties, but not intentional properties, are epiphenomenal due to their nonfunctionalisability. However, it may still be true that (Causal Individuation), but not (Alexander's Dictum), fails for intentional properties. For a nonphysicalist defence of this view, see Chalmers [1996, pp. 150-65].

²³ Contrast with jade: the property of being either jadeite or nephrite. See also Fodor [1997].

²⁴ Kim [op. cit].

²⁵ There are a number of similarities between emergentism and nonreductive physicalism. Both endorse the physical monist claim that all concrete particulars are physical, and both hold the view that mental properties are irreducible yet require physical bases that in some sense are sufficient for them. Horgan [1993, pp. 344-8] takes the difference between them to be that only emergentists deny (Completeness), and that only (nonreductive) physicalists claim that (Supervenience) is in principle explicable within physics itself. Crane [2001b, pp. 207-24], however, believes the only significant distinction between them is epistemological. To be sure, if the nonreductive physicalist blames the argument on (Completeness), there is no significant ontological difference between them. I argue in Sec. IV that there is a better way out of the argument, hence that they are committed to distinct ontological claims.

²⁶ Mellor [1995, pp. 101-105] thinks there are real yet rare cases of overdetermination. He agrees that they are more palatable if there is some causal dependency between the overdetermining causes, for instance if C and C* have a common cause, or there is a

causal chain from C through C* to E. With such examples in mind, he argues that there is no swift inference from the implausibility of overdetermination to physicalism.

²⁷ I assume the standard Lewisian semantics [1973a]: $A \Box \rightarrow B$ is true iff there are no worlds at which A is true, or B is true at all of the closest worlds - the worlds most similar to the actual world - at which A is true, where world-similarity is evaluated in terms of conformity to natural laws and match of particular facts.

²⁸ Loewer (2002, p. 318) calls it ‘dependent overdetermination’. See also Menzies (2001) and Bennett (2003). Thanks to an anonymous referee for clarification on this point.

²⁹ See his [1993, p. 252]. Kim [2003, p. 162] now seems to have accepted this point.

³⁰ For similar views see Horgan [1997], Loewer [2001], Menzies [2001], Yablo [1992].

³¹ Kim [2003, p. 163] has recently expressed the belief that (Exclusion) “is virtually an analytic truth with not much content”. I find this puzzling. A being sufficient for B is consistent with C being sufficient for B, so what is it about the concept of causation that makes it a priori impossible that some effects have two distinct sufficient causes?

³² Crane [1995, p. 219].

³³ For a clear statement see Crane [2001a, p. 61]. Loewer [2001, pp. 46-7] holds that no physicalist can allow “for mental properties to make a causal difference over and above causation by physical properties”, so presumably $\sim M \Box \rightarrow \sim P^*$ is true in virtue of the truth of $\sim M \Box \rightarrow \sim P$ and $\sim P \Box \rightarrow \sim P^*$. Note that they are only committed to the claim that counterfactuals are not always intransitive. We know enough counterexamples to be suspicious of the claim that counterfactuals are always transitive.

³⁴ Contrast with Kim’s [1993, pp. 92-108] ‘supervenient causation’: the causal relation between M and M* supervenes on the causal relation between P and P* in virtue of the supervenience of M on P, and M* on P*, respectively. This account does not meet our desideratum, because the former relation is a mere epiphenomenon of the latter relation. Similarly, Jackson and Pettit [1990] have suggested that mental properties are causally relevant by ‘programming for’ their physical effects, whereas only physical properties are causally efficacious by ‘producing’ their physical effects.

³⁵ One objection concerns effects of a common cause. Suppose C and X are both effects of a common cause O, and that C, but not X, causes E. Then it would seem to be true that

$\sim X \Box \rightarrow \sim E$, hence that X ought to be a cause of E . The truth of this counterfactual follows by transitivity from $\sim X \Box \rightarrow \sim O$, and $\sim O \Box \rightarrow \sim C$, and $\sim C \Box \rightarrow \sim E$. One response is to dispute the backtracking counterfactual $\sim X \Box \rightarrow \sim O$.

³⁶ See also Yablo [1992, pp. 277-8]. Note that $\sim P \Box \rightarrow \sim P^*$ is true only if it is assumed that the closest not- P worlds are worlds in which if M occurs, M has an alternative base property that does not cause P^* . But if $M \ \& \ \sim P \ \& \ P^*$ worlds are closer than $M \ \& \ \sim P \ \& \ \sim P^*$ worlds, the closest not- P worlds better be ones in which M does not occur at all.

³⁷ Some will say this argument equivocates on the way that the relevant worlds are described: the first is true only if they are considered as actual, but the second is true only if they are considered as counterfactual; see for instance Chalmers [1996, p. 133].

³⁸ This modification of his earlier theory [1973b] accounts for pre-emption. Suzy throws a rock at a bottle, and so does Bill. Bill's rock gets there first and so pre-empts Suzy's rock with respect to the shattering of the bottle. So, had Bill not thrown his rock, the effect would have happened anyway, thanks to Suzy's rock. The solution is roughly that there is only a causal chain of actual events leading from Bill's rock to the effect. This, however, does not explain trumping where both candidate causes run to completion, e.g. a major and a sergeant at once shouting 'advance' at the soldiers. Lewis accommodates these cases by introducing a pattern of causal dependence: whether/how/when the effect occurs depends counterfactually on whether/how/when the cause occurs. Note that this implies that in our execution example neither shot is a cause of the convict's death, but as intuitions about such cases of symmetric overdetermination are unclear, this may not be a difficulty for his view.

³⁹ Kim [1998, p. 70] asks for some metaphysical reasons why $\sim M \Box \rightarrow \sim P^*$ is causal. The realisation view may do the job. Assuming M is a second-order functional property, having M is having one of its realiser properties. Having M causing P^* is therefore, as I said in Sec. III, having P causing P^* when an instance of P realises M . M is a cause of P^* iff (i) M has a realiser property P , and (ii) P causes P^* . See also Kim [1993, pp. 326, 362-5; 1999, p. 16].

⁴⁰ A promising account is Menzies' (1996; 2001): C causes E in situation s iff the counterfactual dependence between C and E picks out a real process that is actual in s .

On his view, the existence of neurophysiological processes from brain-states to intentional behaviour do not rule out the existence of psychological processes from reasons to such behaviour, since each process is picked out by the relevant counterfactuals within distinct models.

⁴¹ Thanks to Ned Block, Tim Crane, Lars Gundersen, Jakob Hohwy, Klemens Kappel, and an anonymous referee for very helpful comments.