



Philosophical Review

Mental Causation without Downward Causation

Author(s): John Gibbons

Reviewed work(s):

Source: *The Philosophical Review*, Vol. 115, No. 1 (Jan., 2006), pp. 79-103

Published by: [Duke University Press](#) on behalf of [Philosophical Review](#)

Stable URL: <http://www.jstor.org/stable/20446882>

Accessed: 28/11/2011 18:15

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



Duke University Press and Philosophical Review are collaborating with JSTOR to digitize, preserve and extend access to *The Philosophical Review*.

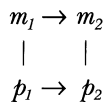
<http://www.jstor.org>

Mental Causation without Downward Causation

John Gibbons

University of Nebraska, Lincoln

The problem of causal exclusion is that an intuitive response to an intuitive picture leads to counterintuitive results.¹ Suppose a mental event, m_1 , causes another mental event, m_2 . Unless the mental and the physical are completely independent, there will be a physical event in your brain or your body or the physical world as a whole that *underlies* this event. The mental event occurs at least partly in virtue of the physical event's occurring. And the same goes for m_2 and p_2 . Let's not worry about what exactly 'underlying' or 'in virtue of' means here.² Here's the picture.



The horizontal arrows represent causation, and the vertical lines represent underlying, whatever that may be. There's some reason to think that the only way m_1 can bring about m_2 is by bringing about p_2 .

I would like to thank Ned Block, Jaegwon Kim, Jennifer McKittrick, Joe Mendola, Steve Yablo, and two anonymous referees for the *Philosophical Review* for helpful comments on previous versions or ancestors of this article.

1. Jaegwon Kim has presented this problem in a number of places. See, for example, *Mind in a Physical World* (Cambridge, MA: MIT Press, 1998). Page references in parentheses are to this work.

2. In this article, I must set aside the important issue of exactly what relation of realization or dependence holds between the mental and the physical. Nothing I say will depend on the details of this relation.

You can't convince someone of something through mental telepathy. You need to interact with the physical world, perhaps by saying something and so making some noise, or by pointing and getting them to turn their head and see. What goes for the case of two people goes for the case of one person as well. Superstition aside, there is no purely mental energy that floats free of the merely physical workings of the brain.

If m_1 brings about m_2 by bringing about p_2 , then m_1 brings about p_2 . This is downward causation. But wait. Doesn't p_1 bring about p_2 ? Isn't that what the bottom arrow represents? Maybe m_1 and p_1 work together to bring about p_2 . There are little holes in the physical causal structure that need to be filled by mental events. You don't need a sweeping metaphysical thesis about the causal closure of the physical to find this implausible. Maybe p_2 is overdetermined. Maybe, but it's not clear what this means. Maybe the idea is this. If m_1 occurred without p_1 , some other physical event would have occurred that would both underlie m_1 and bring about p_2 . But this is still a picture of one physical event bringing about another while the mental gets a free ride. Maybe the idea is that m_1 by itself, in the absence of any physical event, could have brought about p_2 . This holds in ordinary cases of overdetermination. But then you do have a hole in the physical causal structure filled by a mental event, perhaps not in the actual world, but in a nearby possible world.

If m_1 can only bring about m_2 by bringing about p_2 , and if m_1 can't bring about p_2 if downward causation is impossible, then mental causation is impossible. No step in the argument is completely beyond suspicion. But we do have an intuitive reaction to an intuitive picture that leads to counterintuitive consequences. Something here needs to be fixed.

The cure for the problem of causal exclusion, as that problem has been presented, is a token-identity theory.³ Every mental event is a physical event, so each causes what the other does. The events, m_1 and p_1 , are not independent causes, either in the sense of partial or overdetermining causes. They're the same thing. This is true as far as it goes, but the problem immediately reappears at the level of types. Suppose one event causes another and that each of these events has both a mental property and a physical property. Now the picture looks like this.

3. See Donald Davidson, "Mental Events," in *Essays on Actions and Events* (Oxford: Oxford University Press, 1980), 207–27.

Mental Causation without Downward Causation

$$\begin{array}{ccc} M_1 & & M_2 \\ & c \rightarrow e & \\ P_1 & & P_2 \end{array}$$

Was the fact that the cause had a certain mental property, M_1 , causally responsible⁴ for the fact that the effect had a physical property, P_2 ? This would be downward causation at the level of types. If M_1 was causally relevant to P_2 , what about P_1 ? If both M_1 and P_1 are causally relevant to P_2 , we need some account of the relation between these claims of causal relevance. Understanding these on the model of partial causes or on the model of overdetermination is as problematic here as it was in the case of tokens. And in the case of types, the identity theory is even more problematic.

Some versions of nonreductive materialism are noncompetitive.⁵ For some relations between facts or features, the relata do not compete for causal relevance. Identity is clearly a noncompetitive relation in this sense. If x is y , then x 's causal relevance on an occasion does not threaten to preempt or exclude y 's causal relevance on that occasion. In fact, in the case of identity, x 's relevance guarantees y 's relevance. Perhaps there

4. The relation of causal responsibility or causal relevance is a relation between types analogous to causation, which is a relation between tokens. I will also talk about the causal relevance of facts. When the fact consists in something having a property or a pair standing in a relation, you can take this as talking about the causal relevance of the relevant properties and relations. The relation between causal relevance and causal explanation is much more complicated since the latter notion has both a metaphysical side and an epistemic or pragmatic side. At least most of the time, causal explanations work by pointing to a causally relevant property, but what counts as "pointing to" will be different in different contexts as a result of differences in interests and background knowledge. By choosing simple examples and providing relevant information, I will try to manipulate the context so that being explicit is the contextually relevant way of pointing to causally relevant properties.

5. See, for example, Stephen Yablo, "Mental Causation," *Philosophical Review* 101 (1992): 245–80. Yablo thinks that the relation between the mental and the physical is the relation between determinable and determinate, and he thinks that it's a "truism" that "determinates do not contend with their determinables for causal influence" (259). Also see Sydney Shoemaker, "Realization and Mental Causation," in *Physicalism and Its Discontents*, ed. Carl Gillett and Barry Loewer (Cambridge: Cambridge University Press, 2001), 74–98. Shoemaker thinks that the physical realizes the mental and that one property realizes another, roughly, when the causal powers bestowed by the latter are a subset of the powers bestowed by the former. Since distinct properties are bestowing the same causal powers on the same occasion, what makes the view noncompetitive is also what gives rise to at least the appearance of overdetermination.

are other noncompetitive relations as well. Some possibilities include the determinable/determinate, part/whole, realization, and supervenience relations. The strategy is to find a noncompetitive relation such that the mental stands in this relation to the physical. Perhaps part of what's behind this strategy is the idea that if the mental did compete with the physical, the mental would lose.

I think that the mental does compete with the physical. I think that determinates compete with determinables, parts compete with wholes, what is realized competes with its realizer, and functional properties compete with the properties that play the roles. I'm not worried about this threat to nonreductive materialism or mental causation because I think that in the relevant cases, the mental properties win. If you want to know which features of the cause are relevant to a mental feature of the effect, it's probably a mental property of the cause if there is one and not its physical properties. But the physical properties of the cause are relevant to the physical properties of the effect.

On this view, there's no real sharing, just dividing the spoils. Anything one gets, the other doesn't. This is not epiphenomenalism about mental kinds since mental properties are causally relevant to other mental properties. In fact, since the mental properties of the cause beat out the physical properties in the competition for causal relevance, this is a view where mental properties have causal powers over and above, or at any rate other than, their physical realizers. Since the mental properties of the cause are not relevant to the physical properties of the effect, we do not violate the causal closure of the physical. This is mental causation without downward causation.

The view looks good from afar. If it works, it may be the best way to fix what's broken in our intuitive reaction to the intuitive picture. One question is whether we can make it work. Another question is whether there's any reason to believe that it's true. The view breaks naturally into two parts. It says that there's no downward causation. This comes from a fairly strong reading of the causal closure of the physical, and we'll look at this in the first section of the article. It also says that there is mental causation. This comes from a natural conception of levels, and we'll look at that in the second section. Finally, there are various threats to trying to have it all, and we will look at some of these in the final four sections.

What's Wrong with Downward Causation?

Here's one way to think about the causal closure of the physical:

(CCP) Every physical event that has a cause has a physical cause.⁶

But compare this with what Jaegwon Kim has to say about closure.

If you pick any physical event and trace out its causal ancestry or posterity, that will never take you outside the physical domain . . . If you reject this principle, you are ipso facto rejecting the in-principle completability of physics . . . Never mind a complete physical explanation of everything there is; there couldn't even be a complete physical explanation of everything physical. (40)

(CCP) does not seem to capture the idea. Suppose a physical event p is massively overdetermined. It has a physical cause that is sufficient given the laws of nature. But it also has an immaterial mental cause as well as a nonphysical magical cause. Perhaps angels and enteleshies get involved as well. This is a physical event that has a cause, and it has a physical cause. So this situation is compatible with (CCP). But to give a complete causal explanation in an overdetermination case, you need to mention all of the overdetermining causes. So there would be no complete physical explanation of our overdetermined physical event. So (CCP) does not guarantee the in-principle completability of physics.

Of course, our world is not like my imagined case, at least, not according to the physicalist, nonreductive or otherwise. We think that the physical level forms a closed, comprehensive system.⁷ This is why there are strict or exceptionless laws at the physical level if there are any strict laws at all. This is one of the things that make the physical level special. If that's what you think, you should accept what we might call the complete causal closure of the physical: only physical events cause physical events. There are no causal intrusions from outside the physical domain, not even redundant causal intrusions. Since I am primarily concerned with causation at the level of types, or the causal relevance of mental and physical properties, we can understand the complete causal closure of the physical like this:

6. See, for example, Kim, *Mind in a Physical World*, 38.

7. Davidson, "Mental Events."

(CCCP) Only physical kinds are causally relevant to the physical properties of physical effects.⁸

Why should we believe that the physical is completely closed in this way? We can rule out on empirical grounds any kind of mental-to-physical downward causation that involves actually making a difference. Downward causation that made a difference would be nonredundant causation. It would do something that wasn't already going to happen anyway. So the mental would have to be able to violate the laws of physics, or the laws of physics would have to be different inside and outside of brains, or there would have to be new fundamental physical forces that only appear in brains. There's a fairly straightforward way to test these empirical hypotheses. Simply construct MRIs, EEGs, and CAT scans using the physical principles that hold outside of brains. If the devices work as expected on brains, this is strong empirical evidence that there are no new forces that make a physical difference in or around brains. I'm no brain surgeon, but I think that's the way the evidence goes.

But perhaps there's another kind of downward causation that doesn't involve actually making a difference. Perhaps when one event occurs in virtue of another, or one fact obtains in virtue of another, or one property is exemplified in virtue of another, the dependent entity inherits the causal powers of the entity it depends on.⁹ So suppose Smith fatally shoots Jones. Smith's shooting or the fact that Smith shot Jones is causally relevant to Jones's death. But Smith or Wesson shot Jones in virtue of the fact that Smith shot Jones. So maybe the fact that Smith or Wesson shot Jones is causally relevant, maybe even just as relevant as the fact that Smith did. And someone who weighs 150 pounds shot Jones in virtue of the fact that Smith did, so maybe this fact is just as relevant as the others. And maybe the extent to which what you want makes a difference to what you do is the same as the extent to which the shooter's

8. This is stated partly in terms of "kinds" by which I mean natural kinds. I am fairly liberal when it comes to properties. I believe in the property of *being either red or round*. I just don't think that property is a natural kind because it is not a causal power or a causally relevant property. For reasons to believe in many properties while retaining a distinction between natural and unnatural kinds, see David Lewis, "New Work for a Theory of Universals," *Australasian Journal of Philosophy* 61 (1983): 343–77.

9. This is similar to what Kim calls "the causal inheritance principle" though used for the opposite purpose. Kim is concerned with the idea that the realized cannot have more power than the realizer. But someone who is less concerned about overdetermination than Kim might think that if the realized inherits the power of the realizer, this is power enough.

weight makes a difference to the victim's death. Merely redundant mental causation may be better than nothing. But this option is clearly the last resort. Let's see if we can do better.

The ban on mental-to-physical downward causation does not follow from (CCCP) alone. For that, you need the nonreductive part of nonreductive materialism:

(NR) Mental kinds are not physical kinds.

I believe (NR) on the basis of considerations about multiple realization.¹⁰ If there can be two physically quite different individuals, perhaps a Martian and a robot, who both believe that *p*, it doesn't look as though the similarity between them consists in falling under a physical kind. But if there is an important similarity between them, if they do fall under some natural kind or another in virtue of believing that *p*, then that mental kind is not a physical kind. There may be other reasons for believing (NR), and the multiple realization argument is not without its problems.¹¹ But we can simply assume (NR) for the sake of argument. If (NR) is false, there is no problem of causal exclusion since identity is a noncompetitive relation.

Putting our two premises together, we get the argument against mental-to-physical downward causation:

(CCCP) Only physical kinds are causally relevant to the physical properties of physical effects.

(NR) Mental kinds are not physical kinds.

(NDC) Mental kinds are not causally relevant to the physical properties of physical effects.

This conclusion does not entail epiphenomenalism since it says nothing about the possibility of mental-to-mental causation. This argument does not rule out the possibility of downward causation in general. It is specifically about the physical level. Since I think that the multiple-realization considerations are as plausible for geological, biological, social, and economic kinds as they are for mental kinds, I think that

10. For multiple realization, see Hilary Putnam, "The Nature of Mental States," in *Philosophy of Mind*, ed. David Chalmers (Oxford: Oxford University Press, 2002), 73–79.

11. See, for example, Jaegwon Kim, "Multiple Realization and the Metaphysics of Reduction," in *Supervenience and Mind* (Cambridge: Cambridge University Press, 1993), 309–35.

there are true premises analogous to (NR) for these other cases. But I don't think that the analogue of (CCCP) holds at the higher levels. To borrow a line from Donald Davidson, too much happens to affect the mental [and the geological, biological, social, and economic] that is not itself a systematic part of the mental [and so forth].¹² Higher-level laws have exceptions because those levels are not causally closed.

So our argument does not rule out the possibility of downward causation from the social to the mental. Perhaps there's something inherently wrong with the very idea of downward causation, and we need to rule it out one way or another. I don't know if this is necessary or desirable; I just don't think that closure is the way to do it. The argument also fails to rule out the possibility of upward causation, and this is probably a good thing. Though in one respect, the argument is essentially about the physical, in another respect, the argument does generalize. If geological, biological, social, and economic kinds are not physical kinds, there is a causal exclusion problem for them as well. If the mental and friends are not causally relevant to the physical, we need to consider the possibility that there may still be intralevel causal relevance.

What's a Level?

There are many levels, for example, the physical, chemical, geological, biological, psychological, and social levels. There may be many levels even if each level reduces to those below it.¹³ But if the higher levels do not reduce, we may have some reason to take them more seriously. If we can make sense of higher-level causation without downward causation, we have a way to deal with the general threat of epiphenomenalism posed by the problem of causal exclusion. But what is a level?

Levels are natural kinds of natural kinds.¹⁴ The basic idea here is that of a family of properties, and parenthood has an important role to play in the generation of families. Parenthood is a minor modification of a relation defined by Nelson Goodman.¹⁵ The property of *having seven*

12. Davidson, "Mental Events," 224.

13. This was the view of Paul Oppenheim and Hilary Putnam in "Unity of Science as a Working Hypothesis," *Minnesota Studies in the Philosophy of Science* 2 (1958): 3–36.

14. For the full story on levels, see my "Levels" (unpublished manuscript available on author's Web site, www.unl.edu/philosop/people/faculty/gibbons/gibbons.html).

15. "A predicate 'P' is a parent of a given predicate 'Q' if among the classes that 'P' applies to is the extension of 'Q.'" Nelson Goodman, *Fact, Fiction, and Forecast* (Indianapolis, IN: Bobbs-Merrill, 1965), 104. So for Goodman, the extension of 'shade of red'

grams of mass, a property of particulars, is a natural kind. It is a determinate of the property of *having some mass or another*. But the property of *being an amount of mass*, a property of properties, is a parent of the property of *having seven grams of mass*. *Amount of mass* is a kind of kinds, and *seven grams of mass* is one of those kinds. A first-order property is a property of individuals; a second-order property is a property of first-order properties; and when a higher-order property is had by a lower-order property, the former is a parent of the latter.

The amounts of mass form a family. They don't form a family because they have the same last name ('... grams of mass'). They have the same last name because they form a family. Similarities between the amounts of mass are not determined by linguistic stipulation or convention. Similarities between the amounts of mass are as real or objective as the similarities between those particulars that all have seven grams of mass. This is natural realism, or realism about natural kinds. When we're talking about natural kinds, in addition to the objectivity of the similarities, we also expect that the similarities will be causal.¹⁶

If the law says that $f = ma$, then it quantifies over all amounts of mass. It treats all amounts of mass the same way. You don't have one law for big things and another law for little things. You just have one law and infinitely many possible instances. Any causal law of nature that quantifies over amounts of mass expresses exactly the kind of objective, causal similarity that we expect from natural kinds. The standard physical magnitudes, for example, temperature, velocity, distance, and so on are all families of kinds in this way.

In addition to these immediate families, there are also extended families. Mass, acceleration, distance, and so on are all physical magnitudes. They are physical kinds. Copper, gold, salt, and water are all chemical kinds. Different chemical kinds have different degrees of thermal and electrical conductivities and different boiling and melting points. But chemical kinds *have* degrees of thermal and electrical conductivities as well as boiling and melting points. Biological, psychologi-

is a set of sets, and the extension of 'scarlet' is one of those sets. But really, the extension of 'shade of red' is not a set of sets; it's a set of properties, and *scarlet* is one of those properties. This is the minor modification.

16. For the causal individuation of kinds, see Jerry Fodor, *Psychosemantics* (Cambridge, MA: MIT Press, 1988), chap. 2; and Sydney Shoemaker, "Causality and Properties," in *Identity, Cause, and Mind* (Cambridge: Cambridge University Press, 1984), 206–33.

cal, and economic kinds do not generally have melting points. There's something objective and causal that different chemical kinds have in common. So *being chemical* is a natural kind of natural kinds. So chemistry forms a level.

If there are psychological laws that quantify over contents (If you believe that p and desire that q . . .), this is good reason to believe not only that *being a belief that snow is white* is a natural kind, but also that the various kinds of beliefs (beliefs with different contents) form a family. If there's something both causal and objective in the relevant sense that the various propositional attitudes have in common, then psychology forms a level. This is, of course, an empirical question. What is and isn't a natural kind is as empirical as a question can get. As a practicing folk psychologist, I can assure you the empirical evidence I've seen is overwhelmingly in favor of psychology forming a level.

If you think about levels in terms of kinds, laws, or causal similarities, then you shouldn't think about levels in terms of the part/whole relation or the micro-macro hierarchy. What makes galaxies do what they do is pretty much the same as what makes the planets do what they do, even though planets are parts of galaxies. And it's pretty much the same thing that makes rocks do what they do, at least when it comes to rolling down hills. You don't have one law for big things and another law for little things. It's the same law because the same kind of causal powers are at work.

So there are different conceptions of levels. How do we choose among them? I think the choice is easy. If you are at all moved by the multiple-realization considerations, you should leave open the possibility that there are individuals who are on the same psychological level as us but who are not composed of the same biological or chemical parts as us. You should also leave open the possibility that there are very, very large thinkers and, perhaps, very, very small ones. If sameness of level were determined by what you're made of or how big you are, all of these possibilities would be ruled out.

If levels are natural kinds of natural kinds, then same-level causation is the norm. When it comes to natural kinds, not just any similarity will do. Similarities that make a causal difference constitute natural kinds. Mass, charge, position, and velocity are all on the same (physical) level because changes in these features are systematically, causally responsible for changes in the other kinds that constitute the extended family. If psychology forms a level, that will be because what you know, see, think, or want makes a systematic, causal difference to what you

conclude, decide, or do on purpose. Any family of kinds that is systematically, causally relevant to the mental is itself mental. If systematic causal relevance is what ties the immediate families together into the extended family of a level, we do not thereby rule out all cases of upward or downward causation. But we do rule out systematic upward or downward causation.

What's the Problem?

Mental causation without downward causation is just one instance of the more general pattern of intralevel causation without interlevel causation. On the fairly natural conception of levels that I've sketched, what gives unity to a science is the more objective causal unity of the subject matter. It's not just that these events resemble those events by sharing their causally relevant features. It's that these causal transactions resemble those causal transactions by involving the same *kind* of causally relevant features. So if this causal transaction resembles other chemical causal transactions, it does so because the chemical properties of the cause are causally relevant to the chemical properties of the effect. And if the very same causal transaction also resembles other neurological transactions, it does so because the neural properties of the cause are causally relevant to the neural properties of the effect. And if it resembles mental transactions, mental properties are causally relevant to mental properties. What could possibly be wrong with this view?

According to Kim, mental-to-mental causation requires downward causation. If he's right, I'm not. Kim provides two different arguments for his claim (42–43). First, suppose that the mental supervenes on the physical. In terms of our picture, suppose that the mental property of the effect, M_2 , supervenes on the physical property of the effect, P_2 . If we say that c 's being M_1 is causally relevant to e 's being M_2 , and we say that M_2 supervenes on P_2 , then we have two different explanations of e 's being M_2 . M_2 is there because M_1 caused it. And M_2 is there because it supervenes on P_2 . According to Kim, there's a tension between these two explanations, and the only way to reconcile them is to say that M_1 caused M_2 by causing P_2 . But this is downward causation.

I must admit, I don't see the tension between the two explanations. One of them is causal, and the other is not. When you have two distinct causal explanations of a single fact, event, or feature, I do think that there is an apparent tension between them. Perhaps in some cases, both explanations make reference to partial causes. Perhaps in some

cases, you've stumbled upon an instance of genuine overdetermination. But if these options are ruled out on a particular occasion, then there is competition for causal relevance. But if one of the explanations is not causal, it's hard to see how it competes for causal relevance.

Kim also has a more direct argument for his claim. Consider the following principle: "To cause a supervenient property to be instantiated, you must cause its base property (or one of its base properties) to be instantiated" (42). Though the principle does not make reference to causing one thing by causing another, that idea is not far away. As Kim says, you can't directly make a painting more beautiful. You make the painting more beautiful by changing the physical properties on which its beauty supervenes. This is the kind of example that Kim uses to support his principle, and this is a case of making one change by making another. If we accept the idea that you can't make a mental change except by making a physical change, and we also accept that mental facts can't make physical changes, we're faced with the familiar difficulty.

I think Kim's principle, along with the associated idea of making one change by making another, best fits the case where you're thinking about the cause as a token or a particular. They are much less plausible when you're thinking about the causal relevance of various properties. Let's begin with an ordinary case of action. You can't turn on a light directly. You have to flip a switch, clap your hands, or something like that. So suppose you turn on a light by flipping a switch. What caused the light to go on? You did. What caused the change in the switch? You did. And you, of course, are a particular, not a type, kind, or feature. This is where it's most natural to say that you can't make one kind of change except by making another.

If you're like me, you think that your turning on the light on this occasion just is your flipping the switch.¹⁷ We have two different descriptions of the same action, or, more relevantly, two different features of the same effect. When we turn from asking about which thing caused the event to looking for features of the cause that explain the different features of the effect, we get a different picture. "Why did you flip the switch?" "In order to turn on the light." "Why did you turn on the light?"

17. If you're not like me, and you think that events are property exemplifications, then you think that for every different property, there's a different event. So you individuate events pretty much the same way I individuate facts or events-having-properties. If we individuate them in the same way, it won't matter much for present purposes what we call them. For events as property exemplifications, see Jaegwon Kim, "Events as Property Exemplifications," in *Supervenience and Mind*, 33–52.

Mental Causation without Downward Causation

“In order to see better.” When you do one thing by doing another, you can have different explanations for different descriptions of the same event. You can have different explanations for different descriptions because different properties of the cause are relevant to different properties of the effect.

Your total mental state (*c*) causes the action (*e*). Your wanting to see better is relevant to your turning on the light. Your wanting to turn on the light is relevant to your flipping the switch. And your turning on the light depends on your flipping the switch.

wanting to see better	$c \rightarrow e$	turning on the light
wanting to turn on the light		flipping the switch

Is your desire to see better relevant to your turning on the light because it's relevant to your flipping the switch? No. In fact, it's just the other way around. Your desire to see (plus your belief that you can do so by turning on the light) is reason enough to turn on the light. That's what it makes sense to do given these mental states. But these mental states are not reason enough to flip the switch. That's why we need another feature of your mental state to explain this feature of the effect. To the extent that your desire to see better is relevant to your flipping the switch, it's relevant to flipping the switch because it's relevant to turning on the light.

The case of ordinary action explanation provides a useful model for the general case of mental causation without downward causation. Most of the crucial features are here. Different features of the cause are relevant to different features of the effect. There's causal competition without sharing or systematic overdetermination. There are even dependence relations among various features of the events. It does all take place on a single level, so we don't have all of the features. But it is only a model. The model is useful because when causes are reasons, it's easy to see which features of the cause go with which features of the effect. When all goes well, you explain an intentional action in terms of the reasons that make sense of it.

If there's nothing metaphysically wrong with the causal structure I've identified in the case of intentional action, if we can't rule out that causal structure on the basis of general metaphysical or causal principles, then those principles can't rule out the causal structure in the more general case of intralevel causation without interlevel causation. If we understand Kim's principle as talking about what an agent must cause, or when you must cause one thing by causing another, then it is not incompatible

with the causal structure. If we understand the principle as talking about the causal relevance of properties, then it says that a feature of the cause can only be relevant to a dependent feature of the effect by being relevant to the feature of the effect it depends on. The case of intentional action is a counterexample to this claim. There's no plausibility to the idea that you can only be a reason for the end by being a reason for the means. If anything, you're a reason for the means by being a reason for the end, even though the end depends on the means.¹⁸

What about Causal Inheritance?

Kim has one further general metaphysical or causal principle that threatens my view. This is the Causal Inheritance Principle. Here's one formulation.

(CIP) If mental property *M* is realized in a system at *t* in virtue of physical realization base *P*, the causal powers of *this instance of M* are identical with the causal powers of *P*.¹⁹

There are other formulations in other places,²⁰ but I think that the basic idea is this. If some particular mental fact or feature, this thing's being *M*, ontologically depends on (is realized by, supervenes on . . .) some particular physical fact or feature, the thing's being *P*, then the dependent entity cannot have more causal powers than the entity it depends on. After all, where would they come from? Kim says that denying (CIP) means believing that causal powers magically emerge at the higher levels and that this leads to difficulties with downward causation and the causal closure of the physical. Of course, if the magic powers were only exercised at the higher level, we might not have these other problems. But magical emergence is bad enough.

The mental depends on, supervenes on, and is realized by the physical. Mental properties compete with their physical realizers for

18. Does it matter that turning on the light doesn't supervene on flipping the switch? No. Suppose the beauty of the painting supervenes on the total application of paint to canvas. "Why did you put just that paint just there?" "In order to make it beautiful." "And why did you make it beautiful?" "So that the dopes at the gallery will buy it." Your attitude toward the dopes explains making it beautiful but not the particular application of paint.

19. Kim, "Multiple Realization and the Metaphysics of Reduction," 18. The emphasis is in the original.

20. The formulation in *Mind in a Physical World* (54) is in terms of functional properties and the properties that realize them.

causal relevance. At least sometimes, with respect to the mental properties of the effects, the mental properties win. Mental properties of the cause are responsible for mental properties of the effects, and the physical realizers are not. So the dependent instance of the mental property has causal powers distinct from the physical realization it depends on. So I must deny (CIP). Does this depend on a belief in magic? No. But it does depend on a certain view of causation or causal relevance. It depends on taking generality seriously.

It wasn't the bolt's giving way that caused the bridge to collapse. It was the bolt's giving way so suddenly.²¹ And it wasn't the bolt's giving way suddenly and loudly that caused the collapse. It would have collapsed whether the snap was loud or not. Some people think we have three events in the same place at the same time because the different events have different effects. I think this is a job for the causal relevance of properties. But let's set aside the individuation of events. The intuitions behind the argument are intuitions about causal competition. This didn't cause it; that did. Whether this and that are events, features, or facts, they are intimately related, and if the intimacy involves dependence, we have a counterexample to (CIP).

The bolt's giving way supervenes on the bolt's giving way suddenly.²² And the bolt's giving way suddenly supervenes on the bolt's giving way suddenly and loudly. The relevant properties, whether they are properties of the bolt or properties of events, line up nicely in the determinate/determinable hierarchy. If this is how you think of realization, dependence, or determination, you should think that the bolt's giving way suddenly depends on the bolt's giving way suddenly and loudly. But if you agree that it was the bolt's giving way suddenly and not its giving way suddenly and loudly that caused the collapse, then we have a case where the dependent entity has a causal power that it does not inherit from the entity it depends on. This is a counterexample to (CIP).

Perhaps you think the dependence relation goes in the other direction.²³ The bolt's giving way suddenly depends on the bolt's giv-

21. The example originally comes from Davidson, "Causal Relations," in *Essays on Actions and Events*, 161.

22. When we're talking about fine-grained events, one event supervenes on another when the constitutive property of the one supervenes on the constitutive property of the other.

23. The pioneers of fine-grained events, Kim and Alvin Goldman, did not think of dependence, or "level generation" as they called its converse, in terms of supervenience. Xantippe's becoming a widow depends on or is the result of Socrates' death,

ing way. Since the former and not the latter caused the collapse, we still have an example where the dependent is more powerful than what it depends on. Examples abound, and nothing turns on the direction of the dependence. Suppose that Sophie the pigeon is conditioned to peck at red triangles and that she pecks at one that happens to be scarlet.²⁴ It wasn't the fact that the triangle was scarlet that caused the pecking. It was the fact that it was red. If this instance of redness is realized by or depends on the triangle's being scarlet, this is a counterexample to (CIP). If you think the dependence runs in the other direction, train Sophie to peck at scarlet.

To see the examples as counterexamples to (CIP), you need two things. You need to read the examples competitively. Though your favorite dependence relation may obtain between the relevant properties, each is still in a position to screen off, preempt, or exclude the other's causal powers. If you say that both the red and the scarlet are causally relevant to Sophie's pecking, you need some account of the relation between these causal claims. Examples of partial causes and overdetermination do not provide a useful model for this case.

In addition to reading the examples competitively, you need to think that the more general, determinable, supervening property wins the competition. It's fairly clear that if there is a competition, red wins. If it were an empirical question, one quick application of Mill's Method of Difference would pretty much settle the issue. Find a triangle that's red but not scarlet and see what happens. But the philosophical questions are why red wins and how it manages to do so without magical emergence. Red wins without relying on magic because it is at the right level of generality. It's not too specific; it's not too general; it is, as Stephen Yablo says, "proportional" to the effect.²⁵

but the death, while sufficient in the circumstances, isn't really a supervenience base. This is the standard case. If you turn on a light by flipping a switch or vote by raising your arm, we have dependence without supervenience. For Goldman, John's saying 'hello' loudly is generated by, and so depends on, John's saying 'hello'. But saying 'hello' loudly doesn't supervene on saying 'hello'. Saying 'hello' supervenes on saying 'hello' loudly. For Goldman on level generation, see *A Theory of Human Action* (Englewood Cliffs, NJ: Prentice-Hall, 1970), chap. 2.

24. The example comes from Yablo, "Mental Causation," 257, though I'm not sure he would approve of the use to which I'm putting it. For me, this is a paradigm case of competition between determinables and determinates.

25. Yablo, "Mental Causation," 277.

What's So Great about Generality?

One of the biggest mysteries about nonreductive materialism is the idea that a property could be both determined and a causal power. How could it both be that something makes it the case that you're *F* and also that your being *F* makes it the case that you have certain causal powers? Why doesn't the something, whatever it is, that makes it the case that you're *F* also make it the case that you have the causal powers purportedly conferred by being *F*? In fact, how could something make it the case that you're *F* without making it the case that you have those causal powers? But if it's really something else that confers the causal powers, what work is left for *F* to do?

The best answer to this puzzle has to do with taking generality seriously.²⁶ If you and I were both made *F* in different ways, then any account that only made reference to the *F*-makers, our differences, would fail to see what we have in common. You want some ice cream and can see that there's some right in front of you. All you have to do is reach out and grab. You are also, at the same time, in an enormously complex physical state that realizes not only your desire for and perception of ice cream, but also your physical ability to move your body in certain ways. Anyone in just that physical state would make the same bodily movements that you do. This is an important causal fact, and it leads to a genuine worry. Given that the physical state was sufficient for the movement of your arm in the direction of the ice cream, what further work is left for the desire to do?

But there is another important causal fact. Other people who know what you know and want what you want will do what you do, even if they are in a different complex physical state and even if their doing it involves different bodily movements. If we only made reference to the specific physical state you're in before you reach for ice cream, we miss the generalizations that cover both you and those other people. We fail to see what you and the others have in common. So far, so good. But why should we care about those generalizations? Why should we care about what you have in common with them? If we are only interested in these generalizations because they are useful, or predictive, or easy to verify, in short, if the virtues of the generalizations are epistemic or pragmatic, then it could still turn out that as far as genuine causation is concerned,

26. Jerry Fodor has been emphasizing the importance of generalizations from the very beginning. See "Special Sciences," in Chalmers, *Philosophy of Mind*, 126–35.

there's physics and then there's stamp collecting. Doing higher-level sciences may be a nice way to pass the time, but these higher-level sciences are not genuinely concerned with what makes things happen.

What is the objective, metaphysical, or causal significance of generality? The intuitive notion of proportionality, or the right degree of generality, is built into the notion of a natural kind, at least if you think of natural kinds in terms of causal powers or in terms of conferring causal powers. But there are different ways of thinking about kinds as causal powers. We start with the idea that if *being F* is a natural kind, there's something causal that the *F*s have in common. There are different ways of understanding this basic idea. Suppose the *F*s do A, or they cause *Es*. This is something causal that the *F*s have in common. But suppose that different *F*'s do A for different reasons. Suppose that you and I are both glass breakers. We're both disposed to break glass. But I throw bricks through windows because I'm angry at the world, while you break your used jars because you think that makes them easier to recycle. There's something causal we have in common. There's something we both do. But there's no one thing in virtue of which we both do it.

There is a slightly stronger conception of the causal individuation of kinds. If *being F* is a natural kind, it's not enough that *F*s do A. Something has to make *F*s do A. And it has to be the same thing that makes different *F*s do A. The same property, *being F*, must be causally relevant on the different occasions. This conception of kinds leads to pressures in two directions. Any category, or property, or kind divides the world into two piles. There are the things inside the category and the things that are not. If your category is too big, the things inside the category may do the same thing, but they won't do it for the same reason, or in virtue of the same causally relevant feature. This was the problem with *being a glass breaker*. If your category is too small, there will be things outside the category that not only behave in the same way as the things inside the category but that do the same things for the same reason. There may be some difference between things inside and things outside the category. But that difference won't be a causal difference.

The basic idea behind the push toward generality is the idea of causal insensitivity. Sophie can't tell the difference between different waves of the same color light or slightly different shades of red. Perhaps when we talk about Sophie, our talk about telling the difference is not wholly metaphorical. But when we say that the tree can't tell the difference between spring and a temporary thaw in the middle of winter, all we mean to say is that the tree is causally sensitive to some things, the current

Mental Causation without Downward Causation

temperature, and causally insensitive to other things, next week's weather. If something is causally insensitive to the presence or absence of a certain feature, then the presence of that feature cannot causally explain and is not causally relevant to any of the further features of the thing.

Sometimes, instead of complete causal insensitivity, we have causal insensitivity in certain respects. If African emeralds cost more than other emeralds, not all features of emeralds are causally insensitive to their point of origin. Where they come from has a causal influence on their price. But other features of emeralds, their geological features, may be causally insensitive to point of origin. Properties like color, hardness, and structure may be causally sensitive to temperature, pressure, and the presence of impurities. But the fact that these conditions obtained in Africa rather than somewhere else seems like a paradigm case of a causally irrelevant feature. If the difference between African and non-African emeralds makes no geological difference, then the more specific property, *being an African emerald* is not a geological kind even though the more general property, *being an emerald* is a geological kind.

The difference between the complex physical state that underlies my perception of and desire for ice cream and the complex physical state that underlies your perception of and desire for ice cream makes no mental difference. We both do the same thing for the same reasons. These different types of state may be physical kinds, and they may have different physical effects, but they are not mental kinds. If the differences make no mental difference, you can't explain a mental feature of the effect in terms of one of these complex physical kinds. If the same mental property is realized in two different ways, but the mental effects are causally insensitive to these differences in realization, the more general mental property is causally relevant while the more specific property is not. The more general, higher-level properties of the causes are responsible for the more general, higher-level properties of the effects, and the lower-level properties are not because the higher-level properties of the effects are causally insensitive to the specific details of what's going on down there. This is upper-level causation without magical emergence.

What about Headaches, Perception, and Action?

So maybe mental causation without downward causation does not rely on magic. Maybe, there are no counterexample-free, general metaphysical principles that rule out the very possibility of the relevant causal structure. But who needs general metaphysical principles when you've

got everyday cases of interlevel causation? You get a headache and take an aspirin. The aspirin makes the headache go away. Isn't this a case of physical-to-mental causation? You see that there's a book on the table. Doesn't the fact that *p* cause the belief that *p*? You want some ice cream so you go to the store. Isn't your desire causally relevant to the change in location of your body, and isn't that downward causation? If the desire for ice cream can't get you any, how have we saved mental causation?

Why would I want to deny that these are cases of upward or downward causal relevance? There is a certain kind of causal relevance, systematic causal relevance, that ties families of kinds together to form a level. It's not enough that *C*, the property of the cause, is at the same level of generality as *E*, the property of the effect. *C* and *E* must belong to families where other members of *C*'s family are at the same level of generality as other members of *E*'s family. Since this relation puts things on the same level, it cannot hold across levels. If these cases of upward and downward causation are not instances of this kind of causal relevance, then this part of my view is happy to admit them. Even without an analysis of the relevant kind of relevance, I think it's clear that these cases do not count. But I also have an argument against the possibility of mental-to-physical downward causation. If what we want makes a difference to what we do, then either there's something wrong with that argument, or the mental causation of action does not involve downward causation.

Let me take the objections in reverse order. The case of action is the most troubling since it is the only one that purports to involve downward causation. But this case is also the easiest to deal with. If you're only concerned about token causation, the token-identity theory handles that. Of course the desire caused the change in location, as did the physical state that realized the desire since they are one and the same thing. But what about causal relevance? What was the feature of the cause in virtue of which it produced an effect of a certain sort, say, the complex bodily movements that constituted your trip to the store? It looks as though some complex physical or neurological property of the cause is responsible for producing an effect of this sort. People in similar mental states who live closer to the store will produce different bodily movements in order to get there. But someone in the exact same physical state will produce the same bodily movements regardless of whether there's a store at the end of those movements or not.

If the physical property of the cause is responsible for the effect's being of a specific bodily movement type, what further work is left for the mental property to do? On my view, if mental kinds are systemati-

cally, causally responsible for different types of behavior, there should be a psychological family of behavioral types that are at the same level of generality as the mental kinds that explain them. Is there such a family of behavioral types? Of course there is: things you do on purpose.

Sid breaks a window in order to break in to someone's house. Nancy breaks a window because she's trying to throw something through what she thought was an opened window. Lightning strikes a tree. It falls over and breaks a window. Sid, Nancy, and the tree all break a window. Sid intentionally breaks a window, but Nancy and the tree do not. Nancy breaks the window in the course of trying to do something else, but the tree does not. There are similarities and differences among the behaviors. Sid's reasons are causally responsible for his behaving in a certain way. But what way of behaving, or what feature of his behavior are they responsible for? If the right degree of generality matters, then it's not a feature that his behavior shares with Nancy's and the tree's. It's not *breaking a window*.

You can't give the same kind of explanation of Sid breaking a window and the tree breaking a window. There may be something that window breakers have in common, but that something is not a feature on the mental level. You can't give the same kind of explanation of Sid breaking a window and Nancy breaking a window either. Good or bad, Sid acts on reasons for breaking a window. Whether you intentionally A depends on the causal relevance of your reasons for A-ing. Nancy, unlike the tree, acts on reasons for something. But they're not reasons for breaking a window. If Nancy intentionally throws something, then we can give the same kind of explanation of her throwing that we give of Sid's breaking. The agent has some goal and thinks that A-ing will lead to that goal, and so on.

You have a level not just when different causal transactions involve the same causally relevant properties. You have a level when different causal transactions involve the same kind of causally relevant properties. When it comes to the psychological explanation of behavior, bodily movements don't matter; intrinsic duplication doesn't matter; what matters is what you do on purpose. So ordinary action does not involve downward causation. Mental properties of the cause are responsible for the mental properties of the effect. Mental properties are causally relevant to behavior because *intentionally breaking the window* is a mental property, and it's the kind of kind that does the work in psychological explanations.

Sometimes things happen for a reason. When we explain Sid's breaking a window in terms of his goals, we say that this causal process

resembled, in a fairly general respect, Nancy's throwing something and your going to the store for ice cream. The kinds that fit into this pattern are not only intentional actions but certain changes in view and some acquisitions of desire, those changes and acquisitions that are the results of reasons. Taking this pattern seriously is seeing the mental as a level and assigning causal relevance to mental properties.

So if our reasons are responsible for what we do on purpose, does this mean there's something wrong with explaining your location in terms of your desires? Should we have people stop doing this? The topic of explanation is more complicated than the topic of causal relevance. The success of an explanation depends on many things, including the interests and background knowledge of the people involved. But explanation also has a metaphysical side. When explaining is a matter of fitting a particular into a pattern, the patterns have to be there. So of course you can explain people's locations in terms of their beliefs and desires. It's just that when you think about it, you see that part of what makes the explanation work is the fact that they went there on purpose. Then it's up to your audience to get from the fact that they went there to the fact that they're there.

So what about cases of perception? Doesn't the fact that p cause the belief that p ? Though I don't rule out all cases of upward causation, I do rule out systematic upward causation since systematic causal relevance puts things on the same level. Isn't the fact that p causing the belief that p systematic enough? No, as a matter of fact, it's not. Of course, the fact that p is a partial cause of the belief that p . No one should deny this. So it's not as though the fact that p has nothing causal to do with the belief that p . But what was it about the fact that p that made it cause the corresponding belief? If you're looking for something that different cases of perception have in common, it's not merely that the resulting belief happens to be true.

It's just not true in general that if p , then you'll believe that p . Whether or not you notice the fact that p depends on how interesting, obvious, and surprising that fact is. It also depends on what you already know and what you want to know, among other things. Who should we hire to find out what sorts of things make a systematic difference to what you notice? This looks like a job for a psychologist rather than a geologist or botanist. Of course, it's a psychological question which kinds are psychological kinds. So it's a psychological question which questions are psychological questions. But if our initial hunch is correct, then what's systematically connected to your noticing that p is not merely the fact

that p is true. It's the fact that p is interesting, obvious, or what have you. These certainly look like psychological kinds. You couldn't make reference to something's being interesting, obvious, or surprising in your definition if you were trying to give a reductive analysis of the mental.²⁷

We turn, finally, to that annoying headache. We could give a response to this case that fits the mold of the responses to the previous cases. We would need a causal power or a natural kind that is plausibly on the mental level and that is at the right level of generality to beat out *taking an aspirin* in the competition for causal relevance. *Taking a pain reliever* is an obvious candidate. If you had taken a pain reliever that wasn't an aspirin, your headache would still have gone away. And if you had taken an aspirin that wasn't a pain reliever, perhaps because you're a Martian and aspirin doesn't help your headaches, your headache wouldn't have gone away. Of course, I don't think that there's something that it's like to be a pain reliever. But I don't think of being mental in terms of the intrinsic nature of your qualia. I think of being mental in terms of playing a systematic causal role with respect to other mental kinds. But on my view, your playing that role is incompatible with having someone else play that role for you.

The problem with *taking a pain reliever* is not that it's not mental enough. It's that it might not be a causal power. *Being a pain reliever* may be a true functional kind. There may be nothing that all pain relievers have in common except having some other property that relieves pain. But if that other property is relieving the pain, and there's no systematic overdetermination, it doesn't look as though *being a pain reliever* really adds anything. Different pain relievers may do the same thing. But they don't do it for the same reason. If *being a pain reliever* is a functional kind in this sense, then it's epiphenomenal.

If you're not worried about the epiphenomenality of functional kinds, this route is open to you. Since I am, it's not open to me. What other options are there? Perhaps there's nothing systematic to be said about curing pain per se. Perhaps it will turn out that aspirin doesn't work on Martians, that horse tranquilizers don't do to us what they do

27. But wait. The relevant causal power may be psychologically contaminated, but it still entails that p is true. Can't we tell a priori that nothing like that could be a psychological kind? No. That's not the kind of thing you can tell a priori. But more importantly, it's only when philosophers are talking about psychology that the existence of the external world is supposed to be irrelevant. When psychologists are doing psychology, our interaction with the world is actually the point. See, for example, the nearest psychology textbook.

to horses, and that what cures pain isn't really a psychological question because it all depends on the neurophysiology. If you treat two cases of pain differently, not because they differ in some mental respect, but because they differ in some neurophysiological respect, then you're not responding to what the cases have in common. You're not responding to the mental property of *being in pain*.

Perhaps there is something fairly general to be said about relieving human pain. Given similar brains and bodies, we expect the same drug to do similar things. But these expectations are based on physical, not mental, similarities. Where these similarities end, our expectations about the effectiveness of our drugs end as well.²⁸ But even the complete story on how to cure human pain will not involve us in systematic upward causation. *Human pain* or *pain realized by physical state P* is no more a psychological kind than *African emerald* is a geological kind. The difference between human and Martian pain might make some kind of difference, but it doesn't make a mental difference.

Does this mean that pain isn't really a mental kind? No. Pain is a mental kind because being in pain is a good reason to believe you're in pain. It's a good reason to avoid similar situations in the future, whether you're a human or a Martian. And it's a good reason to believe that something bad is going on in your body since that, after all, is how it seems. Pain is not a mental kind because of its intrinsic "ouchiness." It's a mental kind because it makes a systematic mental difference by being a reason.

Part of what's involved in causal relevance is the right degree of generality between *C*, a property of the cause, and *E*, a property of the effect. Sometimes *C* is too specific for *E*. Sometimes *E* is too specific for *C*. But sometimes, it's not just that *C* and *E* are made for each other, that is, they're at the right degree of generality with respect to each other. Sometimes, *C* and *E* are members of families that are made for each other. When you have this, you have a level.

28. I am inclined to take a similar line on cases where a thought causes an increase in your heart rate or it causes you to blush or break out in a cold sweat. These look like cases of downward causation, but considerations of generality suggest that the mental properties may not be doing the work. Although we do expect the connections between kinds of thoughts and these kinds of bodily reactions to hold for other humans, we do not expect them to hold for nonhumans, even those who are capable of having and fully appreciating the relevant thoughts. I would like to thank an anonymous referee for the *Philosophical Review* for raising this objection.

Mental Causation without Downward Causation

When you have same-level causation, the property of the cause is in a position to beat out other contenders in the competition for causal relevance with respect to *E*. Considerations of generality automatically favor the home team. But the very same property of the cause is likely to lose the competition for lower-level and higher-level properties of the effect, and for the same reason. If there's a property of the cause at the same level as the property of the effect, generality will favor it. So, for example, the mental properties of the cause are responsible for the mental properties of the effect, while the physical properties of the cause are responsible for the physical properties of the effect.

This may seem to lead to undue constraints on the scope of psychology. All psychologists really explain is your intentional action on the basis of your knowledge, perception, beliefs, desires, and intentions; your beliefs on the basis of perception and appearance; what you notice on the basis of what's obvious and interesting; change in view and change in plan on the basis of your overall mental state; and all the rest of the goings-on on the mental level. All psychologists really explain are psychological processes. This is the price of the view. But perhaps it's a price we can live with.