

Causation and Mental Causation

Jaegwon Kim

I

An epistolary event occurred in 1643 that will live in the history of the debate on mental causation. In the May of that year, Princess Elisabeth of Bohemia dispatched to Descartes what must be one of the most celebrated philosophical letters, challenging Descartes to explain:

How the mind of a human being can determine the bodily spirits [i.e., the fluids in the nerves, muscles, etc.] in producing voluntary actions, being only a thinking substance. For it appears that all determination of movement is produced by the pushing of the thing being moved, by the manner in which it is pushed by that which moves it, or else by the qualification and figure of the surface of the latter. Contact is required for the first two conditions, and extension for the third. [But] you entirely exclude the latter from the notion you have of the body, and the former seems incompatible with an immaterial things.¹

A need for explanation arises for Elisabeth because she takes contact as a necessary condition for physical causation: the cause—at least, the proximate cause—of the motion of a material body must be in spatial contact with that body, a condition that plainly cannot be met by an immaterial causal agent outside physical space. The idea that causation requires contact survives even in Hume, a philosopher who is commonly thought to have held a deflationary view of causation as consisting solely in de facto regularities. One of the conditions Hume laid down for causation is that of contiguity in “space and time” between cause and effect, either direct or mediated by a chain of contiguous cause-effect pairs. (We will recur to the contiguity condition below.)

¹ Elisabeth to Descartes, May 1643. This quotation is taken from Garber (2001, 172).

Elisabeth's challenge is surely reasonable and intelligible, both in commonsensical terms as well as in light of what Descartes had written, in *Meditation II*, about bodies and causes of their motions:

By a body I understand whatever has determinate shape and a definable location and can occupy a space in such a way as to exclude any other body; it can be perceived by touch, sight, hearing, taste or smell, and can be moved in various ways, not by itself but by whatever else comes into contact with it (Descartes 1641, 17).

There may well have been earlier philosophical concerns about the powers of the mind to bring about changes in the physical world,² but, for many of us, the exchanges between Descartes and Elisabeth are our first encounter with the mental causation debate in the modern period. Descartes continues to loom large in contemporary discussions of many central issues in the philosophy of mind, and our current concerns with mental causation are no exception.

What is of interest to us here is Elisabeth's invocation of a specific feature of mind-body causation in her challenge to Descartes, namely that to cause a material body to move, physical contact with the body is required. Such contact, in modern terms, presumably imparts energy, or transfer momentum, from one body to another, and this fact constitutes the causal relation between them. Elisabeth's complaint, which could be our complaint, is that such a conception of causation leaves no room for mental causation within Descartes's dualism. Minds, being essentially extensionless and not even in physical space, cannot meet the contact requirement; in fact, we can attach no coherent sense to the idea of contact between an immaterial mind and a material body. So either mental causation, at least the mental-to-physical variety, is a metaphysical impossibility, or the very idea has no intelligible sense.

² See Caston (1997).

II

Mental causation has been a flash point of debates in the philosophy of mind for over three decades, ever since the publication of Donald Davidson's "Mental Events" (1970). In this paper Davidson claimed that there are no "strict" laws about mental phenomena—neither mental-physical laws nor mental-mental laws—and that such laws are required to underwrite causal relations. This at first blush appeared to quickly entail that mental causation was not possible. However, Davidson had a deft reply: All that his two principles imply is that any causal relation must instantiate a strict *physical* law, and that what is required for a mental event to enter into a causal relation is for it to have an appropriate physical description under which it can instantiate a physical law. From this Davidson's "anomalous monism" follows: All individual mental events—in fact, all individual events³—that enter into causal relations are physical events.

The early debate on mental causation in the contemporary period began when several philosophers noticed,⁴ apparently independently, that although Davidson's anomalous monism allowed individual (or "token") mental events to be causes and effects, it failed to give any role to mental properties, or mental descriptions, of these events in determining what causal relations they enter into. The reason is simple: since, on Davidson's view, all strict laws are physical and they apply to individual events solely in virtue of the fact that these events have certain physical properties, what mental properties they have—what mental kinds (e.g., pain, desire, thought) they fall under—becomes entirely irrelevant to their causal properties—or so it seemed to his critics. Though Davidson's anomalous monism may not be a form of "token" epiphenomenalism, it was generally perceived as a form of "type" epiphenomenalism (in the terms introduced by McLaughlin 1989), the position that psychological characteristics and features contribute nothing to the causal powers of objects and events that have them. The position has the consequence that if we were to redistribute psychological properties

³ Here we assume that the mental-physical dichotomy is both exhaustive and exclusive. Also, Davidson believes that strict laws (if there are any) can be found only in physics.

⁴ Stoutland (1980), Honderich (1982), Sosa (1984), McLaughlin (1989).

over the events and objects of this world—or even if these properties were entirely removed—that would not change a single causal relation as long as their physical properties are preserved. Davidson tried to defend anomalous monism against the epiphenomenalist charges in Davidson (1993); however, few seem to have found his efforts persuasive.

It did not escape philosophers' attention that Davidson's troubles with mental causation crucially depended on his conception of causation—in particular, the condition that causally related events must instantiate “strict” laws. There has been some controversy about how to understand the strictness of strict laws, or what Davidson meant by the term. Strict laws of course must be laws—that is, as Davidson explicitly notes in Davidson (1970), must be capable of supporting counterfactuals and subjunctives and also be capable of confirmation by observation of positive instances (that is, inductively projectible). There seem two further features that make for the strictness: first, strict laws are totally exceptionless (in this regard, they contrast with laws or generalizations hedged with *ceteris paribus* clauses), and, second, they are often (always?) found as part of a theory that is in some sense “complete” and gives comprehensive coverage over its domain. It is not easy to spell out this second condition in clear terms, something that Davidson himself seems never to have done. In practice, exceptionlessness is what does most of the work, and for most purposes this has seemed sufficient. So a natural question to raise, when we are faced with Davidson's epiphenomenalist predicament, is why we should tie causation to strict laws. Why can't there be causation where there are no strict laws in Davidson's sense? This question is especially appropriate given the fact that Davidson never stated a clear reason, much less a detailed argument, for his requirement of strict laws for causal relations.

III

One strategy that will naturally occur to many is to relax, or abandon, Davidson's requirement of strict laws in favor of a condition that allows nonstrict laws, or laws hedged by *ceteris paribus* clauses, to underwrite causal relations between individual

events. Jerry Fodor is one such philosopher. He admits that “even the best psychological laws are very likely to be hedged” (Fodor 1989, 72), and then continues “it [is] no longer clear why hedged psychological laws can’t ground mental causes; and, presumably, if hedged psychological laws can, then strict physical laws needn’t” (ibid.). But how can a *ceteris paribus* law, a law whose antecedent, say F-events, does not necessitate its consequent, G-events, ground a causal relation between an F-event and a G-event? Given the law, it is amply possible for an F-event to occur without being followed by a G-event. Being qualified by a *ceteris paribus* clause, the law is immune to falsification by such counter-instances; that in fact is the whole point of “*ceteris paribus*” hedges.

Fodor’s reply is based on his special reading of the “*ceteris paribus*” clause. He writes:

The first—and crucial—step in getting what a robust construal of the causal responsibility of mental requires is to square the idea that Ms [mental events of kind M] are nomologically sufficient for Bs [bodily events of kind B] with the fact that psychological laws are hedged... [If] it’s a law that $M \rightarrow B$ *ceteris paribus*, then it follows that you get Bs whenever you get Ms *and* the *ceteris paribus* conditions are satisfied. This shows us how *ceteris paribus* laws can do serious scientific business since it captures the difference between the (substantive) claim that Fs cause Gs *ceteris paribus* and the (empty) claim that Fs cause Gs except when they don’t” (Fodor 1989, 73).

The heart of Fodor’s strategy, then, appears to be the thought that whenever we have a serious *ceteris paribus* law “Ms cause (or are followed by) Bs, *ceteris paribus*”, there is a set C of conditions (as yet not fully specifiable) such that “Whenever C obtains, Ms cause (or are followed by) Bs” is a strict, exceptionless law. The reader will have noticed the alternate formulations, “cause (or are followed by)”, in the preceding sentence. This was to reflect Fodor’s unexplained move, in the quoted passage, from “ $M \rightarrow B$ *ceteris paribus*”, which only states that M is (nomologically) sufficient for B *ceteris paribus* to “Ms cause B (or “Fs cause G”) *ceteris paribus*” (or “Fs cause G”). This slide between

nomological sufficiency and causality occurs throughout Fodor (1989), and it is indicative of the fact that Fodor's conception of causality is strictly and simply based on nomological regularity.⁵ So his problem with *ceteris paribus* laws—that is, his solution to Davidson's epiphenomenalism problem—consists in interpreting “*ceteris paribus*” nomic regularities so that they will yield causation. Fodor's suggestion, then, is that *ceteris paribus* regularities can ground *ceteris paribus* causal claims, and that whenever the unspecified, and unknown, set of additional set of conditions C is satisfied, we have causation *tout court*.⁶

But this, to my mind, is to mislocate the real problem. What one should worry about in this context is not *ceteris paribus* clauses but a more fundamental question about causation and regularities. This question concerns whether or not we can get causation out of regularities, whether these are “strict” or hedged with *ceteris paribus* clauses, or whether they are mere “*de facto*” regularities (Humean “constant conjunctions”) or given a suitable modal force (“physical/nomological necessity”, “lawlikeness”, “projectibility”, and so on). As early as 1925, C.D. Broad made a simple observation (Broad 1925, 96):

“Again, if causation be nothing but regular sequence and concomitance, as some philosophers have held, it is ridiculous to regard psycho-neural parallelism and interaction to be mutually exclusive alternatives. For interaction will mean no more than parallelism, and parallelism will mean no less than interaction.”⁷

Actually, the situation is not be quite as simple as Broad describes, for causation, or causal interaction, has directionality whereas psychoneural correlations under a strict parallelism are entirely symmetric and it would be difficult to determine which of the two symmetrically correlated events, one mental and one physical, is the cause and which the

⁵ Fodor briefly considers the possibility of noncausal laws, that is, nomological regularities that do not constitute causal relations, at Fodor (1989, 65), but waves it off in his typically lighthearted way.

⁶ It's a bit of a mystery how we can ever know these “unknown” conditions are satisfied and hence how we can know a causal relation exists in a given situation.

⁷ Thanks to Brian McLaughlin for bringing this paragraph to my attention.

effect. But that is the least of the problems faced by the regularist, or nomological, approach to causation.

First, there is the much discussed cases in which two phenomena are correlated, with nomological necessity, because they are collateral effects of a single cause.⁸ One of the two effects may always occur a little earlier than the other so that we may mistakenly think that the first is the cause of the second. There must be many such cases in medicine where a single underlying pathological state gives rise to two distinct symptoms, one occurring earlier than the other. The regularity connecting the two symptoms may be projectible and lawlike, and there seems no reason to deny that it holds with nomological necessity (we could even suppose it strict and exceptionless, though this is unlikely). The regularity, though it arises from underlying causal processes, clearly does not constitute a causal relation—a relation in which one event brings about another.

Situations with the following structure present another difficulty. We observe a regular connection between two events, A and B, with A preceding B, and we may be tempted to postulate a causal connection between them, with A as cause and B as effect. In fact, a believer in a purely regularist-nomological conception of causation is committed to this conclusion. However, it may well turn out that the observed correlation between A and B is due to A's regular correlation with event C and B's regular correlation with event D, where C causes D. The correlation from A to B is only a surface manifestation of an underlying causal process involving C and D. It will be easy enough to find instances exemplifying this situation in medicine in which an underlying pathological process gives rise to regular connections between symptoms caused by the various stages in the progression of the pathology. Closer to home, where mental causation is concerned, regular sequences of mental events may well be cases of this kind. If so, the impression that we are observing mental-to-mental causation would only be an impression. The fact would be that the observed sequence of mental events is grounded in, and is explained by, an underlying causal process between the neural

⁸ For extensive discussions of “causal forks”, see Wesley C. Salmon (1984). Broad notes such cases in Broad (1925, esp. 115-117).

substrates of the mental events. There is no direct causal relationship between the successive mental events in the sequence; the only genuine causal relations present here relate the neural substrates of the mental sequence. Thus, the relationship between the two successive mental events is like that between the shadow cast by a moving car at an instant and the shadow of the car an instant later. The moving shadows do not constitute a causal sequence; nor do a sequence of mental events grounded in a series of causally connected neural substrates.⁹ It should be clear that the issues about *ceteris paribus* laws do not touch these difficulties with the regularist-nomological conception of causality.

Trying to soothe our fear of epiphenomenalism (“epiphobia”), Fodor tells us:

According to the present view, the properties projected in the laws of basic science are causally responsible, and so too are the properties projected in the laws of the special sciences. Notice, in particular, that even if the properties that the special sciences talk about are supervenient upon the properties that the basic sciences talk about, that does *not* argue that the properties that the special sciences talk about are epiphenomenal (Fodor 89, 66).

But this is no cure of epiphobia. Our discussion shows that though there may be projectible special-science properties and there may be special-science laws, that does not guarantee that there is causation in the special sciences. Fodor continues as follows:

Not, at least, if there are causal laws of the special sciences. The causal laws of the special sciences and causal laws of basic sciences have in common that they *both* license ascriptions of causal responsibility (Fodor 89, 66).

To be sure, if there are *causal* laws in psychology, they will license ascription of causal responsibility to psychological properties and ground psychological causal relations. The crucial question unaddressed by Fodor is whether psychological laws *are* causal laws—that is, whether the regularities we observe in the psychological domain are causal

⁹ This is the gist of what I have called the “supervenience argument”; see Kim (1998, 2005).

regularities, or mere mirror images of the causal regularities at a more fundamental (presumably, neural/physical) level. Fodor's neglect of this question is a reflection of his seemingly unconscious slide between regularities and causal regularities, and laws and causal laws, which itself seems to be an outcome of his unreflective commitment to a regularist-nomological view of causation.

IV

So the issue of mental causation cannot be resolved either way by merely invoking the regularist or nomological approach to causation. In saying this, I should not be understood as committing myself to the view that the nomological model of causation does not work at any level; it is possible that it works at the fundamental physical level. Because it is the bottom level with nothing below it, regularities may be all we can get; it makes no sense to speak of "underlying" mechanisms or processes at a "lower" level. Or we can perhaps take this to mean that, although only "constant conjunctions", but no causation, exist at the fundamental level (Hume was right about this level), causal relations can, and do, exist at higher levels. These are interesting and intriguing issues but we must set them aside and move on.

The nomological approach, which once was the reigning approach to causation as well as explanation, seems to have lost favor with a large number of philosophers. The tide now seems to have turned in favor of the counterfactual conception of causation whose modern version is due to the seminal work of David Lewis (1973a). This approach has recently attracted a lot of attention from an active and energetic group of philosophers, and there appear to be numerous ongoing research projects attempting to develop a satisfactory counterfactual analysis of causality. The core idea of the counterfactual conception is that *c* is a cause of *e* just in case *e* is counterfactually dependent on *c*—that is, if *c* had not occurred, *e* would not have occurred. Since counterfactual dependence is not transitive whereas causation is, Lewis explained "*c* causes *e*" as the ancestral of counterfactual dependence, that is, in terms of there being a series of events linking *c* with *e* such that any event in this series (except *c*) is

counterfactually dependent on its predecessor. There are numerous outstanding difficulties with the counterfactual approach, among them the problems of overdetermination and preemption—problems that refuse to go away. The current literature is rife with increasingly clever and ingenious counterexamples and equally (if not more) ingenious remedies to evade or dissipate them.¹⁰ The impression one gets looking in from outside is that it is going to be by no means an easy affair to achieve a reasonably simple and intuitively well-motivated account of causation along the counterfactual lines, and the increasing number of epicycles being piled on top of the epicycles already there reminds one of the ultimately fruitless search of the “fourth” condition of knowledge prompted by the Gettier Problem. I don’t want to say that the ongoing research on the counterfactual analysis of causation is without value; far from it, it may yield—I believe it has already yielded—some valuable insights into our causal talk, just as the Gettier-inspired work in epistemology has contributed much to our understanding knowledge and justification. Our present concern is not with the ultimate viability of the counterfactual approach to causation; it is a more restricted one about explaining mental causation in terms of counterfactuals, although in doing this our discussion will unavoidably involve some general issues about causation and counterfactuals.

One such general issue concerns the apparent dependence of counterfactuals, at least those involved in causal attributions, on laws and regularities, and if this is the case, embracing the counterfactual approach to causation will have no advantages over the regularist-nomological approach we considered earlier. Consider the causal claim: The striking of the match caused it to light. On a simple counterfactual analysis, this amounts to the assertion of the following counterfactual:

(C) If this match had not been struck, it would not have lighted.

Almost all current counterfactual theorists of causation use the semantics of counterfactuals developed by Robert Stalnaker (1968) and David Lewis (Lewis 1973b),

¹⁰ An excellent place to sample some of this is Collins, Hall, and Paul (2004).

based on comparative similarity among possible worlds. According to this scheme, (C) is true just in case the consequent of the conditional “the match lighted” is true at the world that, apart from the fact that the match was not struck in that world, is the closest to the actual world in all other respects (we use this somewhat simplified formulation; this will make no difference). Assume (C) is true. That means that in the closest world in which the match was not struck, it did not light. How do we know that this world is closer to the actual world than is the closest world in which the match was not struck but it nonetheless lighted? The obvious, and the only possible, answer seems to be that, in the actual world, dry matches struck in the presence of oxygen usually and reliably ignite, and that it is our knowledge of this regularity, or *ceteris paribus* law, combined with knowledge of the actual circumstances in which the match was struck (e.g., it was dry, oxygen was present, etc.), that accounts for our knowledge of (C). Or the laws involved might be more theoretical and concern the chemical composition of the match head, its combustibility, the characteristics of the surface against which the match was struck, and so forth. In any case, one crucial aspect in which the comparative similarity is judged seems to involve the similarity of laws holding in a given world. In the case of (C), it is difficult to see how evaluations of conditionals like (C) could avoid adverting to laws and regularities.

Let us see how this affects the use of counterfactuals to account for mental causation. Consider the claim that a sudden attack of migraine headache caused Susan a frightful sense of anxiety. For the counterfactualist, this amounts to the truth of:

(D) If Susan had not had the sudden migraine headache, she would not have experienced frightful anxiety.

We can concede that our commonsense “knowledge”, or assumption, that counterfactuals like (D) are often true grounds our belief in the reality of mental causation. Our job as philosophers, in the present context, is to see what makes the likes of (D) true and whether this justifies the claim that Susan’s migraine headache caused her anxiety. If our observations relating to (C) are correct, the truth of (D) must depend on the regularity

connecting sudden attacks of migraine headaches and feelings of anxiety. This regularity could be limited to Susan and a few others like her, or it could be a (*ceteris paribus*) law for all people with migraine headaches. It seems to me that even an epiphenomenalist like T.H. Huxley can, with consistency, accept a regularity of this kind and acknowledge it to be lawlike (surely, the connection isn't accidental or coincidental—not even for an epiphenomenalist). However, the epiphenomenalist will deny that (D) warrants the causal claim that the attack of migraine headache caused the sense of fearful anxiety. The observed regularity arises out of the genuine causal process connecting two neural substrates on which the headache and the anxiety respectively causally depend. The situation is fundamentally the same if you believe that the relationship between mental states and the underlying neural states is better described in terms of supervenience or realization. As we remarked earlier in regard to Fodor's nomological approach, the question is not whether there are these psychological (or psychophysical) regularities or laws; rather, it is whether these regularities are causal regularities and whether these laws are causal laws.

V

In spite of these and possibly other difficulties, the counterfactual approach to causality remains popular—among philosophers working on issues about causation (especially, the analysis of the concept) and among philosophers who aim to defend mental causation against various epiphenomenalist threats.¹¹ The intuition that supports the counterfactual approach, I believe, is the close association we form between a cause of an event and a *sine qua non* condition of its occurrence. A cause is the condition but for which the effect would not have occurred. We can grant the legitimacy of this intuition, without necessarily wedding it to any particular way of making it precise and exact—without, that is, necessarily explicating it in terms of counterfactuals with the Stalnaker/Lewis-style semantics, or any other special semantics of conditionals.

¹¹ For example, LePore and Loewer (1987), Horgan (1989), Loewer (2002).

But there is another strong intuitive conception of causation that contrasts sharply with the conception tied to counterfactual dependency, or the *sine qua non* condition. It is a *productive* or *generative* conception of what causing consists in. On this conception, a cause is something that produces or generates or brings about its effects, something from which the effects derive or spring forth. This idea was given its classic expression when Elizabeth Anscombe wrote:

There is something to observe here, that lies under our noses. It is little attended to, and yet still so obvious as to seem trite. It is this: causality consists in the derivativeness of an effect from its cause. This is the core, the common feature, of causality in its various kinds. Effects derive from, arise out of, come of, their causes. For example, everyone will grant that physical parenthood is a causal relation (Anscombe 1971, 91-92).

Indeed, in a recent article, Ned Hall (2004) makes a plausible case for the thesis that there are two fundamentally distinct notions of causation:

Causation, understood as a relation between events, comes in at least two basic and fundamentally different varieties. One of these, which I call “dependence,” is simply that: counterfactual dependence between wholly distinct events. In this sense, event *c* is a cause of (distinct) event *e* just in case *e* depends on *c*. that is, just in case, had *c* not occurred, *e* would not have occurred. The second variety is rather more difficult to characterize, but we evoke it when we say of event *c* that it helps to *generate* or *bring about* or *produce* another event *e*, and for that reason I call it “production” (Hall 2004, 225).

According to Hall, three characteristics are central to productive/generative causation: transitivity, locality, and intrinsicness. Of these what is relevant to our present concerns is the locality condition, which Hall states as follows: “Causes are connected to their effects via spatiotemporally continuous sequences of causal intermediaries” (Hall 2004, 225).

This is a generalized version of Hume's contiguity condition alluded to earlier. Hume's own statement is this:

I find in the first place, that whatever objects are consider'd as causes or effects, are *contiguous*; and that nothing can operate in a time or place, which is ever so little remov'd from those of its existence. Tho' distant objects may sometimes seem productive of each other, they are commonly found upon examination to be link'd by a chain of causes, which are contiguous among themselves, and to the distant objects; and when in any particular instance we cannot discover this connexion, we still presume it to exist (Hume 1739, 75).

As Hall notes, causal relations conforming to the dependence idea need not meet the locality condition; we will see some examples below.

What we have seen in earlier sections is that neither the nomological nor the dependency conception of causation can properly ground mental causation. I argued that nomological relationships do not deliver the kind of causal efficacy we want for mentality, and that the counterfactual approach seems to presuppose, or collapse to, the nomological conception and thereby inherit the latter's shortcomings. Many counterfactualists will dispute this claim. We need not concern ourselves with this general issue about causation. In this section, I will try to argue that the relation of causation as dependence, or counterfactual dependence, even if it is a proper and useful causal relation, is not the source of our worries about mental causation. That is, even if we succeed in showing that mental causation, with causation construed as dependence, is real, that would not suffice to vindicate mental causal efficacy and thereby dissipate our epiphenomenalist worries. In my view, what we need to show is that mentality has causal powers to bring about their effects in a continuous process of generation and production.

Why should we resort to this "thick" variety of causation in thinking about mental causation? My answer is pretty simple: We care about mental causation because we care about *human agency*, and agency requires the productive/generative conception of

causation.¹² I don't have a knock-down argument to prove that agency requires productive causation; I hope what I will say here makes my claim at least plausible. It seems to me that mere counterfactual dependence is not enough to sustain the causal relation involved in our idea of acting upon the natural course of events and bringing about changes to realize what we desire and intend. An agent is someone who, because she has certain beliefs, desires, emotions, intentions, and the like, has the capacity to perform actions in the physical world—that is, to cause her limbs and other bodily parts (e.g., her vocal cords) move in appropriate ways so as to bring about changes in the natural course of events in the outside world—open a door, pick up the morning paper, and make a cup of coffee. It seems to me that without productive causation, which respects Hall's locality condition or Hume's contiguity condition, such causal processes are not possible. These causal processes all involve *real connectedness* between cause and effect, and the connection is constituted by phenomena like energy flow and momentum transfer, an actual movement of some (conserved) physical quantity.¹³ In saying this we need not impugn the dependency conception of causation; all we need is the point that agency requires productive causation.¹⁴ Note, further, that we need not claim that dependency is not involved in actions; it may well be that the dependency involved, say between a limb movement and a desire, has an explanation in terms of the productive/generative relations between them.

Consider the component of mental-to-physical causation involved in action, namely the causation of bodily movements by our desires, beliefs, intentions, and the like? To endow our mental states with causal powers to move our limbs (or, more proximately, the powers to bring about changes in our neural states), would it be enough

¹² In correspondence Barry Loewer has challenged this claim. According to him, "thin" causation, or dependence, is sufficient to ground agency.

¹³ I am of course referring to the so-called conservative quantity approach to causation. See Dowe (1992, 2000), Salmon (1994); for an early statement, see Fair (1979). See also the exchange between Dowe (2004) and the dependence theorist Schaffer (2004).

¹⁴ For our purposes we need not claim that all cases of action involve productive causation; perhaps we are willing to regard certain cases of omissions as actions and consider omissions as eligible as causes. See below for further discussion of omissions.

to show that counterfactuals like the following are true—and that we have reason to believe them true?

If Susan had not wanted to open the window, neural state N would not have occurred in her brain (where we suppose N triggers an appropriate sequence of bodily movements).

If Susan had not experienced the sudden migraine headache, neural state N* would not have occurred (where we imagine N* to be the neural substrate of an anxiety attack).

We earlier argued how such counterfactuals ultimately involve reference to psychological or psychophysical regularities, and that their significance for mental causation depends on the question whether these regularities are *causal* regularities. Apart from this issue, there are reasons to be suspicious about relying on counterfactuals alone to defend mental causation—what such a strategy could show.

Friends of the counterfactual approach often tout its ability to handle omissions and absences as causes and the productive/generative approach's inability to account for them. We are inclined to take the truth of a counterfactual like:

If Mary had watered my plants, the plants would not have died

as showing that Mary's not watering, an omission, caused the plants' death and take that as a basis for blaming Mary for killing the plants. But obviously there was no flow of energy from Mary to the plants during my absence (that exactly was the problem!); nor was there any other physical connection, or any spatiotemporally contiguous chain of causally connected events.

One problem with regarding Mary's omission as a cause of the plants' death on the strength of the foregoing counterfactual is that there are indefinitely many other

counterfactuals like it certifying an indefinitely large number of other causes of the plants' death:¹⁵

If George W. Bush had watered my plants, the plants would not have died.

If Laura Bush had watered my plants, the plants would not have died.

If Hillary Clinton had watered my plants, the plants would not have died.

Well, you get the idea. Of course I blame Mary for not watering the plants; but we don't have to say that I am blaming Mary for causing the plants to die; I am blaming her for breaking a promise—her promise to water the plants while I was away.

If omissions should count as actions, something we *do*, then by staying in my room “doing nothing” (I could be taking a long nap), I would be performing countless actions, like not watering my plants, not writing an email to my niece, not running the MS marathon, not doing the space walk out of the shuttle Discovery, ... Of course these are not intentional omissions (at least they don't have to be), but it is difficult to see how intentional omissions and mere omissions could differ ontologically, specifically with respect to their causal powers. At any rate, it is by no means clear that its ability to handle omissions as causes is something that the dependency theorists should celebrate. Not thinking, not believing, not desiring, and so on are mental omissions. If causation by mental omissions count as mental causation, that would make mental causation easy—too easy. My not believing (or disbelieving) that a chest of treasures is buried in my backyard is a cause of my not digging in my backyard; my not believing treasures are buried in your backyard causes my not digging in your backyard, and so on ad infinitum. This is not causation worth having.

¹⁵ I saw examples like these in Abbott (1974) for the first time. This paper is recommended to those interested in the counterfactual approach to causation. See also McGrath (2005).

VI

But if we understand causation in mental causation in the productive/generative sense, wouldn't that rule out mental causation—in particular mental-physical causation—too quickly, without any need for an argument? Especially if we require that causation requires energy flow or momentum transfer, how could there be such a process from a mental entity to a physical entity? Remember Elisabeth's challenge to Descartes: the causation of physical motion requires spatial contact, but how could an immaterial mind outside physical space be in such contact with a material body? Notice that this problem is not special to Cartesian physics; it arises even under Hume's concept of causation, which, as we saw, requires a spatially contiguous chain of causally connected events. Don't all such conceptions of causation, conceptions that require some "real" connections between cause and effect, automatically rule out mental-physical causation (and hence human agency)? Further, what could "contiguity" mean unless it meant spatial contiguity? What "real" connection can there be between two immaterial substances outside space? Wouldn't the productive/generative conception of causation preclude, without much ado, mind-to-mind causation as well as mind-to-body causation—that is, all mental causation?¹⁶

An answer—the right answer, in my opinion—is contained in a follow-up letter Elisabeth sent to Descartes. In June 1643, she wrote to Descartes:

And I admit that it would be easier for me to concede matter and extension to the mind than it would be for me to concede the capacity to move a body and be moved by one to an immaterial thing (Garber 2001, 172).

This, I believe, is a remarkable statement which attests to the philosophical astuteness of Elisabeth. She is saying what some of us have been trying to say for the past two decades; she is in effect saying that to make sense of mental causation, she would rather

¹⁶ For an argument for an affirmative answer to this question based on spatial considerations, see Kim (2005), chapter 3.

physicalize the mind (“concede matter and extension to the mind”) than accept the possibility of immaterial minds in causal commerce with material bodies. It is the idea that mental causation is possible only if mentality is physically reducible. Her remark expresses this idea succinctly and forcefully; it may well be the very first causal argument ever for physicalism.

But the dominant strain of physicalism on the contemporary scene has been the nonreductive kind. Nonreductive physicalists, while rejecting Cartesian immaterial minds or any nonphysical object, nonetheless resist the idea that mental properties are reducible to physical properties. Beliefs, desires, intentions, pains, visual images, and the rest, though they may be supervenient on neural/biological processes, are irreducible to them; nonetheless, these mental states are claimed to be causally efficacious. But how can the idea of productive/generative causation be applied to them in relation to neural/physical states? How can there be energy flow or momentum transfer from a desire, as an irreducible mental state, to the firing of a group of neurons? In his characterization of nonreductive physicalism, Terence Horgan, a leading proponent of the position, writes:

First, mental properties and facts are determined by, or supervenient upon, physical properties and facts. Second (and contrary to emergentism), physics is a causally complete science; the only fundamental force-generating properties are physical properties. More specifically, the human body does not instantiate any fundamental force-generating properties other than physical ones. Third, mental properties nonetheless have genuine causal/explanatory efficacy, via the physical properties that “realize” mental properties on particular occasions of instantiation (Horgan 1996, 498).

So, on Horgan’s antireductionist view, a desire (as an individual “token” event) has “genuine” causal efficacy, say the power to raise my arm to reach for a glass of water, in virtue of the fact that its neural realizer, an instance of a neural property on which the desire supervenes, has this causal power.

Horgan's suggestion, I believe, is fundamentally right: mental events and states have the causal efficacy that they have because their neural/physical realizers have causal efficacy. In fact, a mental state, occurring on a given occasion, in virtue of being realized by a certain neural/physical state, has exactly the causal powers of that physical state. Where I differ from Horgan is that once we are prepared to say what we have just said, the next natural step to take—in my view, a step we are compelled to take—is to reductively identify this particular mental state with its neural/physical realizer (Kim, 1993, 1998). This of course is to jettison the “nonreductive” part of nonreductive physicalism. To say what Horgan says, namely that the belief is a distinct state from its neural realizer, and yet consider each a sufficient cause of the arm rising, is to walk smack into the problem of overdetermination. And to say that the mental state has causal efficacy “via” the causal efficacy of its neural realizer carries an apparent epiphenomenal implication: Given that the neural realizer is a full cause of the arm rising, what causal work is left for the mental state to contribute? Or, to put the question another way, what could “via” mean here? What is it for an event to cause something “via” another event that presumably does the real causing?

So the idea of causation as production and generation, or causation as requiring a “real” connection between cause and effect, can be applied to mentality as long as, and probably only so long as, mental states have physical realizers. Whether an approach of this kind leads to reductionism, as I just claimed, or it is compatible with a nonreductive view of mentality is a further, currently much debated, issue (Kim 2005).

References:

- Abbott, Barbara (1974). “Some Problems in Giving an Adequate Mode-Theoretical Account of CAUSE”, *Berkeley Studies in Syntax and Semantics*, vol. 1, ed. Charles Fillmore, George Lakoff, and Robin Lakoff, Berkeley, CA.

- Anscombe, Elizabeth (1971). *Causality and Determination*, Cambridge: Cambridge University Press. Page references are to its reprinted version in *Causation*, ed. Michael Tooley and Ernest Sosa, Oxford: Oxford University Press, 1993.
- Broad, C.D. (1925). *The Mind and Its Place in Nature*, London: Routledge and Kegan Paul.
- Caston, Victor (1997). "Epiphenomenalisms, Ancient and Modern", *Philosophical Review* 106: 309-363.
- Collins, John, Ned Hall, and L.A. Paul, eds. (2004). *Causation and Counterfactuals*, Cambridge: MIT Press.
- Davidson, Donald (1970). "Mental Events", reprinted in Davidson, *Essays on Actions and Events*, New York and Oxford: Oxford University Press, 1980.
- ____ (1993). "Thinking Causes", in *Mental Causation*, ed. John Heil and Alfred Mele, Oxford: Oxford University Press.
- Descartes, René (1641). *Meditations on First Philosophy, The Philosophical Writings of Descartes*, vol. II, tr. John Cottingham, Robert Stoothoff, and Dugald Murdoch, Cambridge: Cambridge University Press, 1985.
- Dowe, Phil (1992). "Wesley Salmon's Process Theory of Causality and the Conserved Quantity Theory", *Philosophy of Science* 59: 195-216.
- ____ (2000). *Physical Causation*, Cambridge: Cambridge University Press.
- ____ (2004). "Causes Are Physically Connected to Their Effects", in *Contemporary Debates in Philosophy of Science*, ed. Christopher Hitchcock, Oxford: Blackwell Publishing.
- Fair, David (1979). "Causation and the Flow of Energy", *Erkenntnis* 14: 219-250.
- Fodor, Jerry A. (1989). "Making Mind Matter More", *Philosophical Topics* 17: 59-79.
- Garber, Daniel (2001). *Descartes Embodied*, Cambridge: Cambridge University Press.
- Hall, Ned (2004). "Two Concepts of Causation", in *Causation and Counterfactuals*, ed. John Collins, Ned Hall, and L.A. Paul, Cambridge: MIT Press.
- Honderich, Ted (1982). "The Argument for Anomalous Monism", *Analysis* 42: 59-64.
- Horgan, Terence (1989). "Mental Quausation", *Philosophical Perspectives* 3: 47-76.

- _____ (1996). "Reduction, Reductionism", in *The Encyclopedia of Philosophy: Supplement*, ed. Donald M. Borchert, New York: Simon & Schuster Macmillan.
- Hume, David (1739). *A Treatise of Human Nature*, London. Page references are to the L.A. Selby Bigge edition, Oxford: Clarendon Press, 1888.
- Huxley, T.H. (1901). *Methods and Results: Essays*, New York: D. Appleton. A useful excerpted version can be found in *Philosophy of Mind: Classical and Contemporary Readings*, ed. David J. Chalmers, Oxford: Oxford University Press, 2002.
- Kim, Jaegwon (1993). "Postscripts on Mental Causation", in Kim, *Supervenience and Mind*, Cambridge: Cambridge University Press.
- _____ (1998). *Mind in a Physical World*, Cambridge: MIT Press.
- _____ (2005). *Physicalism, or Something Near Enough*, Princeton: Princeton University Press.
- LePore, Ernest and Barry Loewer (1987). "Mind Matters", *Journal of Philosophy* 93: 630-642.
- Lewis, David (1973a). "Causation", *Journal of Philosophy* 70: 556-567. Reprinted, with Postscripts, in Lewis, *Philosophical Papers*, vol. II, Oxford: Oxford University Press, 1986.
- _____. (1973b). *Counterfactuals*, Cambridge: Harvard University Press.
- Loewer, Barry (2002). "Comments on Jaegwon Kim's *Mind in a Physical World*", *Philosophy and Phenomenological Research* 65: 655-662.
- McGrath, Sarah (2005). "Causation by Omission: A Dilemma", *Philosophical Studies* 123: 125-148.
- McLaughlin, Brian (1989). "Type Epiphenomenalism, Type Dualism, and the Causal Priority of the Physical", *Philosophical Perspectives* 3: 109-136.
- Salmon, Wesley C. (1984). *Scientific Explanation and the Causal Structure of the World*, Princeton: Princeton University Press.
- _____ (1994). "Causality Without Counterfactuals", *Philosophy of Science* 61: 297-312.
- Schaffer, Jonathan (2004). "Causes Need Not Be Physically Connected to Their Effects: The Case for Negative Causation", in *Contemporary Debates in Philosophy of Science*, ed. Christopher Hitchcock, Oxford: Blackwell Publishing.

- Sosa, Ernest (1984). "Mind-Body Interaction and Supervenient Causation", *Midwest Studies in Philosophy* 9: 271-281.
- Stalnaker, Robert (1968). "A Theory of Conditionals", in *Studies in Logical Theory*, ed. Nicholas Rescher, Oxford: Blackwell.
- Stoutland, Frederick (1980). "Oblique Causation and Reasons for Action", *Synthese* 43: 351-367.