PERGAMON

Studies in History and Philosophy of Modern Physics 33 (2002) 717–724

Studies in History and Philosophy of Modern Physics

www.elsevier.com/locate/shpsb

Essay review

Physical causation

Daniel M. Hausman*

Department of Philosophy, University of Wisconsin, 600 N. Park St., Madison, WI 53706, USA

Phil Dowe, *Physical causation* (Cambridge Studies in Probability, Induction, and Decision Theory), Cambridge University Press, Cambridge, 2000, pp. ix + 224, US$60.00, ISBN: 0-521-78049-7 hbk

Causation is a frustrating subject. Suppose one begins with some promising idea such as that causation is counterfactual dependence or statistical relevance. One then develops this idea with care and intelligence, revises and improves it to cope with criticisms, and by the time one is finished, sane people will be looking elsewhere. If one wants conclusive reasons to reject the counterfactual theory of causation, one can do no better than to read Lewis' (1986) many postscripts. If one wants the best refutation of a probabilistic theory of causation, then one should read my colleague, Ellery Eells' (1991) magisterial defense. In *Physical Causation*, Phil Dowe performs the same service for physical process/interaction theories of causation.[1] Although he does a superb job at clearing away some objections to the theory, by the time Dowe is done, I think it is evident that his conserved quantity theory of causation is not tenable. Causation continues to stump us all.

In saying this, I do not at all mean to imply that Dowe has nothing to teach us. On the contrary, "failures" such as Lewis', Eells', and Dowe's are proud moments for philosophy. In learning just how counterfactual, probabilistic, and process/interaction theories fail, we learn a great deal about causation. What we learn does not constitute a "theory." It is not a set of necessary and sufficient conditions for "*C* causes *E*." But it is knowledge all the same. There are important lessons to learn concerning how causation escapes our grasp, and these lessons are best learned from those who have reached the furthest. Dowe's *Physical Causation* has a great deal to teach everyone who seeks to understand causation.

*Tel.: +1-608-263-3700; fax: +1-608-265-3701.

E-mail address: dhausman@facstaff.wisc.edu

[1] Modesty should not prevent me from noting that I may have achieved as much for the independence theory of causation in my *Causal Asymmetries* (Hausman, 1998).

PII: S1355-2198(02)00038-2

Dowe defends the view that causal processes are space–time trajectories of objects that possess conserved quantities, and that causal interactions are intersections of causal processes in which conserved quantities are exchanged. Conserved quantities are quantities governed by conservation laws. To ''possess'' a conserved quantity q is to instantiate the property of having some amount of q. An uncharged glass rod possesses charge, though the net amount happens to be zero. A beam of light in contrast does not possess even a zero amount of charge. An ''exchange'' of conserved quantities is simply a change in the amount possessed by one process that is (since interactions are governed by conservation laws) matched by changes in amounts possessed by other processes. The notion of an ''object'' carries a heavy weight in the account. The objects that possess the conserved quantities are reidentifiable and must be distinguished from mere ''spatio-temporal junk'' (pp. 91, 98–109). (Here is one place for counterexample-mongerers to congregate.)

These claims of the conserved quantity theory are meant to be general factual claims about the world, like the claims that energy is conserved or that energy is proportional to mass and the square of velocity (pp. 7, 10). Dowe's claims are not meant to analyze either everyday or scientific concepts of causal processes or causal interactions. The fact that this account of causal processes and interactions is inconsistent with a good deal of what people say about causal processes and interactions is, Dowe maintains in Chapter 1, as irrelevant as the inconsistencies between the physical account of energy and everyday claims concerning energy.

Although an ''empirical analysis'' of causation such as Dowe's is clearly a legitimate undertaking, the criteria of adequacy are more complicated than Dowe admits, and the analogy to physical theories, such as theories of energy, is misleading. Physicists might plausibly deny that they have any interest in what non-physicists think of as energy. There is no harm in this, apart from the misunderstanding that may result when neophytes confuse the physicist's E with what people call ''energy.'' But it is difficult for metaphysicians such as Dowe to deny that they are theorizing about what people think of as causation, because there is no esoteric discipline in which a technical notion of ''causation'' plays a role similar to the detailed and precise role that energy plays in physics. Arguments for theories of energy cite results of experiments and the possibilities for conceptual clarification and simplification of physical theories. Dowe's arguments for the conserved quantity theory, in contrast, draw on beliefs concerning causation that enjoy widespread acceptance among both ordinary people and scientists.

Without some plausible connection to what ordinary people and scientists take to be causation, the conserved quantity theory would float free of both physics and philosophy. Unless this theory makes plausible claims about what it is for one thing to cause another, it is not a theory of causation. This is not to say that Dowe cannot sensibly challenge accepted beliefs. For example, he has a strong case for denying that omissions can be causes (Chapter 6). But our everyday commitment to omissions as causes is shaky, and at the same time that he challenges it, Dowe offers a good explanation for why omissions seem so much like causes. But Dowe nevertheless must show that his theory is compatible with a large portion of our most central and secure beliefs about causation. Otherwise he has only provided a relatively uninteresting quasi-physics of "causal" processes and interactions, which has nothing to do with traditional concerns about causation and little to do with physics.

As I have argued before (1998, pp. 14–17), there is a great deal of hard work to do in relating theses concerning causal processes and causal interactions to claims about causation. One might be tempted to say simply that C causes E if and only if there is some set of causal processes and interactions linking C and E. Dowe calls this "the naïve process theory" (p. 146), and he rejects its sufficient condition. For example, my hitting a tennis ball against a wall does not cause the wall to remain standing, even though there is a set of causal processes and interactions linking my hitting the tennis ball to the zero momentum of portions of the wall (p. 149). He argues that some way must be found to avoid finding spurious causal relations everywhere, and in Chapter 7, "Connecting Causes and Effects," he attempts to provide a more sophisticated account.

There is, of course, a much simpler objection to the claim that C causes E if there is a set of causal processes and interactions linking C and E: since "linking," unlike causation, is a symmetrical relation, the sufficient condition is an obvious non-starter. Remarkably enough, Dowe never mentions this problem with naïve process theory; and he only takes up questions concerning the asymmetry of causation in Chapter 8. The best way to understand what he is doing, I think, is to take him as implicitly accepting Mackie's (1980) "factoring" strategy (p. 85; see Hausman, 1998, pp. 55f and Dowe himself, p. 89). On Mackie's view, a theory of causation involves two theoretical tasks. First, one needs a theory of causal connection—that is, of the symmetrical relations that obtain between C and E when C causes E or E causes C[2]—and then one needs a theory of causal asymmetry—that is, of the difference between C causes E and E causes C. In his chapter on "Connecting Causes and Effects," one should take Dowe to be developing a theory of causal connection.

In offering his theory of causal connection, Dowe first canvasses various efforts to marry process and probabilistic theories of causation (pp. 150–67). This discussion is of considerable interest, but it is a digression, since the resulting "integrating solution" does not begin to solve Dowe's problems, and, in any case, Dowe wants a process, not a probabilistic theory. Unfortunately, canvassing probabilistic accounts also contributes to some misapprehensions on Dowe's part concerning what problems he needs to solve. Dowe (mistakenly in my view) seeks to distinguish between (a) Rosen's (1978) case where a squirrel kicks a golf ball headed for the cup yet winds up causing it to go into the cup by a somewhat different route and (b) Nancy Cartwright's (1979, p. 28) case of a plant that survives *despite* being treated with a defoliant, rather than because of the treatment. But both the kick and the defoliant are causally relevant to the outcomes. The difference between the cases is *not* that in one case the intervention is causally connected to the result, while in other it is not. Partly because Dowe does not recognize this, he winds up with an implausible and perhaps even inconsistent theory.

[2] One can also argue (as I have) that effects of a common cause are causally connected to one another, but these questions would take us too far from Dowe's concerns.

Dowe's solution to these difficulties, which he calls the problem of "misconnections" (pp. 146–49), depends crucially on regimenting what can serve as relata of the causal relation. These are, Dowe stipulates, special sorts of events or facts, where events are changes in amounts of conserved quantities possessed by objects and facts consist of objects possessing conserved quantities (pp. 169–70). (This is an idiosyncratic view of facts and of the distinction between facts and events.) For simplicity in presenting his theory, Dowe deals with facts only, and I will follow him here. So the relata of the physical causal relation are all facts such as $q(a) = x$ (with some implicit time reference). This abbreviates the claim that the amount of conserved quantity q possessed by object a at the implicitly specified time is x. Call these relata of the causal relation "physical facts." Physical facts must not be disjunctive or negative (p. 170). Accordingly, $q(a) > x$ and $\sim[q(a) = x]$ are not physical facts, and cannot be the relata of any causal relations. "Manifest" causal relations, that obtain among relata which are not physical facts or events, supervene on causal relations among physical facts and events (p. 172).

Consider just the simplest version of Dowe's account of causal connection:[3] Two interactions are linked by a causal process if the process possesses some conserved quantity q that is exchanged in both of the interactions (p. 171). This view helps with some cases. It may justify the denial that my hitting a tennis ball is causally connected to the tennis ball continuing to be green. But it does not begin to solve the problems. Momentum is exchanged both in my hitting the tennis ball and in the wall continuing to stand, and it is carried by the ball. Little progress has been made with the difficulties that led Dowe to reject the naïve process theory.

Although Dowe is not very explicit about the point, his way of saving the theory depends instead on his account of the relata of the causal relation. Dowe insists that causal connections can only obtain between physical facts or "manifest facts" that supervene on physical facts. Even if a physical fact concerning the momentum of the ball, such as $q(a) = x$, caused a fact concerning the momentum of some portion of the wall just after the impact, such as $q(b) = y$, and $q(b) = y$ entails $q(b) < z$, where z is some critical level that must be reached for the wall to collapse, $q(a) = x$ does *not* cause $q(b) < z$, because $q(b) < z$ is disjunctive and thus cannot be caused (p. 175).

This cure is, I believe, worse than the disease. It means that a great many ordinary, even paradigm causal claims turn out to be false, and it fails to draw the distinctions Dowe hoped to make. Dowe begins his account of causal connection seeking to distinguish between, on the one hand, Rosen's squirrel's kick and Cartwright's defoliant. Perhaps if biologists knew enough about the mechanisms by which the defoliant works, and other scientists had succeeded in reducing chemistry to physics, we might find out that the conserved quantities carried by the defoliant differ from

[3] When more than one conserved quantity is involved, the basic idea of Dowe's account is that $q(a)$ (at t) is causally connected to $q'(b)$ (at t') if there is a set of causal processes and interactions between the two facts in which both q and q' are exchanged (and in which certain nomological conditions are satisfied). So the photon that bombards an unstable atom can be a cause of the subsequent decay, when the decay occurs, because the decay involves the simultaneous exchange of charge and energy. But the photon's absorption cannot be a cause of the atom persisting, "since the effect concerns charge, not energy, and there is no interaction where both energy and charge are exchanged" (p. 173).

the relevant conserved quantities carried by the living plant and that the proper sort of interaction did not obtain. But our judgment about whether the defoliant causes the survival would not be affected by the inquiry, since, according to Dowe's theory, surviving is in any case not the sort of thing that could be a relatum of the causal relation. So the defoliant does not cause the plant to survive—but neither does fertilizing it, watering it, or doing anything else to it. Neither does the squirrel's kick cause the golf ball to fall into the cup, because falling into the cup is disjunctive, too. On the other hand, the defoliant can cause specific physical facts about the still living plant that entails that it is alive, just as the squirrel can cause the golf ball to be moving over the lip of the cup with a particular momentum at a particular time.

Dowe's solution also seems contradictory, or at least obscure. His claim that a fact A may be one of the relata of a causal relation if A supervenes on a physical fact apparently contradicts his claim that A cannot be one of the relata of a causal relation if A is disjunctive. On the view of supervenience that Dowe favors (p. 19–20, 169) if $x < z$, then $q(a) < z$ supervenes on $q(a) = x$. It appears that either Dowe must grant that facts such as $q(a) < z$ can be causal relata, or he must limit the relata of causal relations to (non-disjunctive and non-negative) physical facts. The second alternative is unacceptable, because limiting causal relata to physical facts would imply that almost every causal claim humans have ever made has been false.

A better solution for Dowe would be to bite the bullet, to defend the naïve process theory of causal connection, and to capture the intuition that hitting the tennis ball does not cause the wall to remain standing without denying that there is a causal connection between the two. There is, in fact, nothing implausible about maintaining that there is a causal connection between hitting the ball and the complex physical fact imprecisely described by the claim that the wall is still standing. There are a variety of reasons why we would deny that hitting the tennis ball *caused* the wall to remain standing, but these reasons need not be a part of the account of causal connection. Some are pragmatic, others point to the fact that causality is linked to explanation and that causal language is, consequently, not just a way of pointing out causal interactions and processes.

Before considering the upshot of the difficulties just documented, let us turn to Dowe's theory of causal asymmetry, that is, of the difference between causes and effects. His account begins with the bold premise that quantum mechanics presents us with examples of backward causation (pp. 176f; see also Price 1996, Chapters 8 and 9). In particular, the explanation for EPR correlations is that measurement on one particle in a pair (M_a) causally determines the properties that the pair possessed earlier when they separated (S).

Dowe then considers whether alternative accounts of causal asymmetry can explain this premise. So he can of course immediately rule out temporal priority views. Views that relate causal priority to our perspective as agents are also ruled out, because the direction of the causal link between M_a and S is, Dowe maintains, *objectively* contrary to the usual direction of causation (pp. 189–92). Views that take the causal direction of individual causal relations to match that of the whole causal "net" are unsuitable, because they too rule out backward causation.

Dowe's solution is to maintain that the direction of causation is determined by the direction of open forks, if there are any. Otherwise the direction of causation in the particular instance should match the predominant direction of the whole causal net (p. 204). C, D, and E constitute a causal fork if and only if all three of the pairs are correlated, C is not temporally between D and E, and C screens off the correlation between D and E. A causal fork is open to the future if C precedes D and E and there is no causal fork involving D, E, and any F that succeeds D and E. A causal fork involving C, D, and E is open to the past if D and E precede C, and there is no causal fork involving D, E, and any F that precedes D and E. In order for M_a to cause the earlier event S, there must be a fork that is open toward the past. So there must be an open fork involving M_a, S and some other event or fact x that also precedes M_a. Dowe's commitment to backwards causation, plus his account of causal asymmetry, thus make an empirical prediction: there should be some x, that is correlated with the initial state S of the particle pairs (as later caused and revealed by the measurement M_a), and which correlation is screened off by M_a (pp. 206–7) and by nothing preceding S and x.

Unfortunately, this prediction could be true—this correlation could obtain—only if it were not possible to know whether x occurred until after M_a (or not possible to intervene to change M_a after observing x). Otherwise, after observing x, one could change M_a and thereby show that the apparent correlation between x and S was not a real correlation after all.[4] This is not an argument against the possibility of a fork open to the past. But it must not be possible to tell about x until after the time of M_a. So Dowe's empirical prediction could never be tested. (Note that x could just as well be the measurement on the other particle.)

In my view, this fact eliminates any reason to accept the backward causation story. A crucial feature of explanations that identify causes is that they point to possible "levers" with which to manipulate the explanandum phenomenon. To identify a cause thus has testable implications concerning what will happen when these levers are pushed. Such tests may not be feasible, but when they are not even conceivable, the claim to have identified a cause is empty.

Although these considerations challenge Dowe's premise that there is backwards causation in quantum mechanics, they do not show that anything is wrong with his conclusion—that is, his account of causal asymmetry. The difficulty with his account of causal asymmetry is much simpler. *He doesn't have one.* Both in this book and in earlier articles, Dowe sees the problem of causal asymmetry as finding some indicator permitting one to judge whether a causal process (or some set of processes and interactions) is from A to B rather than from B to A. Some asymmetry has to be found so that, having settled on whether the process is from A to B, we also have a means to settle whether some other causal process is from G to K or from K to G. If

[4] This is, of course, a version of the bilking argument (Dummett, 1964). Dowe mentions that Nick Smith pointed out to him the dangers of bilking, but Dowe takes the problem to be that by preventing M_a, one could bilk the causal relation between M_a and S (p. 207). The real problem, as pointed out in the text, is that one would bilk the correlation between S and x, refute the claim that there is a fork here open to the past, and thereby, on Dowe's account, refute the claim of backward causation.

it turned out that whenever the process goes from A to B, a particular sage and immortal raven croaks, "B is the effect," Dowe's problem would be solved. There would be an objective way to tell.

But in my view, the problem of causal asymmetry would not even have been broached. An account of causal asymmetry should explain why the causal relation is not symmetrical. It needs to say what causal asymmetry or priority *is*, why it *matters* whether A is the cause or the effect. It should explain what the asymmetry has to do with causation rather than citing some adventitious fact that may happen to coincide with the direction of causation. What does the fact that there are few causal forks open to the past have to do with causation? What does the direction in which a fork is open have to do with the fact that causes can be used to manipulate their effects, while effects cannot be used to manipulate their causes? What does the fork asymmetry have to do with asymmetries in explanation or in counterfactual dependence?[5]

Because he does not answer these questions—or even raise them—Dowe has no theory of causal asymmetry. And there is a good reason why not: his account of causal processes and interactions is not relevant to these questions. Any satisfactory theory of causation—if there could be such a thing—would have to contain a large chapter (in my view, much more than half the story) to which the process account has little to contribute.

If one supposes that C causes E if and only if C and E are causally connected and C is causally prior to E (the causal relation runs from C to E rather than vice versa), then the tasks of a theory of causation are to tell us what causal connection and causal priority are. The conserved quantity theory, as we have seen, has nothing to say about what causal priority is, and, as refined by Dowe, it provides no satisfactory account of causal connection. What, then, can we learn from an account of causal processes such as the conserved quantity theory?

Here is one possibility. Perhaps there is no such thing as *the* causal relation. It may be that some of what people believe about causation can be clarified and linked to empirical theorizing by an account of causal processes and interactions such as Dowe's conserved quantity theory. Facts about causal processes and interactions may provide necessary conditions for some causal beliefs and sufficient conditions for others. I suspect that there are, however, other aspects and varieties of causation and that it is futile to offer a theory of some single relation called "causation" in terms of the trajectories of objects possessing and exchanging conserved quantities.

Thanks to Ellery Eells and Elliott Sober for detailed criticisms.

References

Cartwright, N. (1979). Causal laws and effective strategies. *Noûs*, *13*. (Reprinted in *How the laws of physics lie* (pp. 21–34), by Nancy Cartwright, 1983. Oxford: Oxford University Press.)

Dummett, M. (1964). Bringing about the past. *Philosophical Review*, *73*, 338–359.

[5] For an attempt to answer questions such as these, see Hausman (1998).

Eells, E. (1991). *Probabilistic causality*. Cambridge: Cambridge University Press.
Hausman, D. (1998). *Causal asymmetries*. Cambridge: Cambridge University Press.
Lewis, D. (1986). *Philosophical papers*. Vol. II. Cambridge: Cambridge University Press.
Mackie, J. (1980). *The cement of the universe*. Oxford: Oxford University Press.
Price, H. (1996). *Time's arrow and Archimedes' point: New directions for the physics of time*. Oxford: Oxford University Press.
Rosen, D. (1978). In defense of a probabilistic theory of causality. *Philosophy of Science*, *45*, 604–613.