Philosophy Compass 1 (2006) PS 002, 1–9

Philosophy of causation: blind alleys exposed; promising directions highlighted

Ned Hall
Massachusetts Institute of Technology

Abstract

Contemporary philosophical work on causation is a tangled mess of disparate aims, approaches, and accounts. Best to cut through it by means of ruthless but, hopefully, sensible judgments. The ones that follow are designed to sketch the most fruitful avenues for future work.

Contemporary philosophical work on causation is a tangled mess of disparate aims, approaches, and accounts. Best to cut through it by means of ruthless (but, I hope, sensible) judgments. The ones that follow are designed to sketch what I think are the most fruitful avenues for future work.

§1 Causation is not part of fundamental ontology

Some facts about our world are ontologically more basic than other facts. A worthy goal for metaphysics is to delineate the *most* basic such facts – to articulate the fundamental ontological structure of the world, or at least to articulate the most plausible hypotheses about what it might be. In this endeavor, physics is the best guide – not about the gory details, but rather about what, in the abstract, should appear in an inventory of fundamental ontology. Here is one attractive picture, directly inspired by physics.

The fundamental ontological structure of the world comprises three sorts of facts. First, there are facts about the spatiotemporal structure of the world. Second, there are facts about the complete, instantaneous physical states the world is in at each moment of time (or, if you like, along each complete spacelike hypersurface). Third, there are facts about the fundamental laws that govern the evolution of these physical states. (These are laws of the sort that fundamental physics aims to discover, typically formulated by means of differential equations such as Hamilton's equations or the Schrödinger Equation. They are not to be confused with the so-called "laws" of the special sciences, which are really certain kinds of causal generalizations in disguise.)

This picture presents the ontologist with two profoundly different tasks. One is to fully and exactly articulate its elements, and to resolve various controversies to which it naturally gives rise. (Are facts about the

laws redundant, given the other elements – as Lewis (1983) and other "Humeans" believe? How much of spatiotemporal structure should be viewed as metaphysically primitive? And so on.) The other task is to show how other facts – *non*-basic facts, facts *not* directly about the fundamental ontological structure of the world – reduce to basic facts. For reduce they do: that is the force of the word "fundamental".

Only the former task genuinely concerns ontology. An example, to make the point clear: Imagine two philosophers of biology, disagreeing about which things are alive (one thinks viruses are; the other thinks they are not). Now, this *could* be a disagreement at the level of fundamental ontology – e.g., if one of the philosophers maintained vitalism, and the other denied it. More likely, though, the dispute will involve no disagreement whatsoever about fundamental ontology, and consequently there is a real question about what genuine issue these philosophers could possibly be arguing over.

Set that question aside for the moment; we will return to it in the next section. For now, the crucial point to appreciate is that philosophical disputes about the nature of *causation* are like this dispute about life, in that they have – or should have – nothing to do with the fundamental ontological structure of the world. Rather, the proper goal for a philosophical account of causation is to show how causal facts reduce to the more ontologically basic facts about this structure.

Of course some philosophers will dispute this claim. Both Tooley (1990) and Armstrong (2004), for example, argue for a kind of causal primitivism, according to which God could create a world, endow it with a determinate spatiotemporal structure, a complete history of instantaneous physical states, and fundamental laws governing their evolution – and still have work left over, namely, specifying what causes what. But the arguments in favor of this position are tissue thin, turning on outré thought experiments of no probative value (Tooley), or too quick a despair at the prospects for a successful reductive account of causation (Armstrong). And the arguments against are rock-solid. In particular, primitivism generates a skepticism about our knowledge of causal facts for which the usual appeals to inference to the best explanation provide no cure (for the relevant explanatory work is already done by the fundamental laws), and makes an unacceptable mystery out of what should be a straightforward supervenience relation: namely, the supervenience of causal facts at the macrophysical level on causal facts at the microphysical level (see Hall forthcoming). Let us set this view aside: honest toil, after all, isn't as bad as all that.

§2 A philosophical account of causation succeeds not by virtue of snugly fitting the intuitive "data", but by virtue of the utility of the concept or concepts it produces

Return to our philosophers of biology, and the question left open about their dispute over the living/non-living distinction. What could this dispute – reasonably – be about? The best answer is this: It is a certain

kind of *practical* dispute, about what is the most useful way, of the various available ways, of drawing this distinction. These philosophers agree, we may stipulate, about all the relevant biochemical facts. They are not vitalists, and so agree that facts about what is alive should be reduced to these biochemical facts (en route to being reduced to even more ontologically basic facts). But they disagree about the correct form of the reduction. Now, they might foot-stampingly insist that the *correct* reduction is the one that gets the facts about what is and is not alive *right*. But that would be misleading, and at any rate silly. For they are already in perfect agreement about all the more ontologically basic facts that could possibly be relevant. It is not as if the world, having presented us with the biochemical facts, offers up a residual mystery: Where, in and among them, are the facts about life to be found? If anything, what we are left with is a kind of decision about how best to draw the distinction between living and non-living things – how to draw it, that is, in a way that will most usefully serve our theoretical purposes.

Unfortunately, a long tradition in philosophy stands in the way of appreciating the value of this simple methodological point. That is the tradition that says that philosophers should focus attention on concepts of central philosophical interest, and pursue analyses of them by gathering up intuitions about hypothetical cases, gathering up a priori "platitudes" involving the given concept, and seeking a philosophical account of the concept that does the best job of systematizing all of this "data". Happily, there is some evidence that this picture has begun to lose its grip. Philosophers of biology, for example, have for some time now pursued analyses of locutions such as "the function of organ X in creature Y is to **Φ**" that advert directly to the evolutionary history of that organ in the species to which Y belongs. When someone comes along and points out that these accounts of biological function assign no function whatsoever to the "organs" of swamp-creatures (creatures that spring into being as the result of massively coincidental interactions of molecules in a swamp), they rightly shrug their shoulders in dismissal. Rightly, because their analyses earn their keep not by virtue of the way they fit the intuitive "data", but by virtue of their utility in explicating biological practice.

Look just a little ways outside the borders of philosophy, and you will see that the precedents have in fact been around for a long time. A very stubborn "ordinary intuition" about size of collections or sets is the following: if set A is a proper subset of set B, then B must be larger in size than A. For more than a century now, mathematicians have learned to love a concept of size of set (spelled out in terms of one-to-one mappings) that dismisses this intuition out of hand. It is laughable to suggest that this lack of fit with ordinary intuition counts as even a minor blemish on the standard mathematical account of set size.

Such a healthily opportunistic attitude needs to, but has not quite yet, infect philosophical work on causation. Does that mean that we should

just ignore ordinary intuitions about causation, effectively abandoning a methodology that has been the industry standard for the last 30 years? Not exactly. Admittedly, were we in the enviable position of, say, contemporary mathematics, possessed of a rich and exact understanding of the theoretical roles for which a concept or concepts of causation were needed, then we could bid farewell to the intuition-based, proposal-and-counterexample methodology so long in vogue. We can hope that, at the end of the day, we will occupy just such a position. But it's the beginning of the day, and consequently a sensible strategy is to search for precise analyses of causal concepts that look like they might prove useful – useful in, for example, causal theories of X (for any philosophically interesting X), as well as in other areas (see below). Intuitions, both about the causal structures of hypothetical cases and about the general principles governing causation, can be enormously helpful in such a search – not because they constitute bedrock "data" to which an analysis must conform on pain of refutation, but because they offer hints, clues, signposts to where interesting and useful causal concepts might be found. If intuition delivers firm verdicts about a range of hypothetical cases, then that is some evidence – defeasible, but to be taken seriously all the same – that it is latching onto distinctions that a successful analysis ought also to respect. If intuition treats certain general claims about causation – e.g., that it is transitive, so that if C is a cause of D, and D of E, then C is thereby a cause of E – as a priori "platitudes", then that is some evidence – defeasible, but to be taken seriously all the same – that a successful analysis should incorporate these claims. But in the final analysis (as it were), it is the utility of the concepts so produced that matters.

And there are limits on which intuitions deserve respect. "What caused Socrates' death? Lots of things, perhaps – but certainly not his birth!" Does this widespread intuition refute any analysis (that would be most of them) that says that Socrates' birth is among the causes of his death? Of course not. A certain amount of deviation from intuitions such as this ought to be permitted, at least if it can be backed up by theoretically well-motivated argument (as, in this case, it certainly can: see for example Lewis (2004)). In fact, before we even start the project of coming up with an analysis of causation, we can discern some general reasons for suspecting that many ordinary intuitions will be irrelevant. For ordinary usage of causal locutions is often very strongly governed by considerations of salience. Billy strikes a match, thereby lighting it; asked about the causes of the lighting, ordinary intuition naturally hits upon the striking of the match, while ignoring such things as the presence of oxygen in the room. That an analysis of causation draws no such distinction should not matter in the slightest.

One can go further. The literature is rife with examples of what I call "abnormal" causation: backwards causation, causation at a temporal distance, causation under indeterminism. A philosopher of causation is well

within her rights to *start out* by ignoring all such cases, seeking an analysis or analyses that work cleanly under the assumption that the fundamental laws are deterministic, and permit neither backwards causation nor causation at a temporal distance. Suppose she succeeds, wildly: would it make sense to complain at her inability to handle a range of substantially more esoteric cases? No. It *would* make sense to explore whether and how her account might be *extended* to such cases (especially indeterministic cases, which are far less esoteric than the others). But even if it can't, that should prompt a reaction of interested surprise, and not the absurd judgment that her account must have been off-track all along.

An emphasis on theoretical utility as the prime desideratum for an account of causation has one more wholesome effect worth mentioning: It renders it unsurprising that there be *more than one* causal concept worth articulating. There are many jobs, both in philosophy and elsewhere, for which we would like the services of a precisely articulated concept of causation; why suppose that just *one* such concept will suffice? Only close investigation will tell – and in my view, it tells decisively in favor of pluralism (see Hall 2004, and the next section).

§3 Causal relations derive from abstract nomological relations

Consider some bad accounts of causation:

Crude sufficient condition account: C causes E iff C and E both occur, and in any world in which C occurs, and which has the same laws as our own, E occurs.

Crude necessary condition account: C causes E iff C and E both occur, and in any world in which E occurs, and which has the same laws as our own, C occurs.

Simple counterfactual account: C causes E iff C and E both occur, and had C not occurred, E would not have occurred.

Bad, to be sure (though the third improves dramatically on the first two). But they get something important exactly right: all three attempt to analyze causation in terms of an abstract nomological relation – "abstract" in that the only contribution the laws make to specifying this relation is to fix a distinction between worlds that are and worlds that are not nomologically possible relative to our own. (That takes some work to bring out, in the case of the counterfactual account; I won't pause over this point here.) That's the right general approach, although one should not expect the analysis to be so direct: more plausibly, a successful analysis will *build upon* some such simple, clean, abstract nomological relation. We'll see examples shortly.

Meanwhile, let's quickly dismiss two rival approaches that give laws a much more concrete role to play. Regularity approaches try to analyze causal connections as instances of laws (Hume 1748), sometimes relative to particular ways of describing the cause and effect (Davidson 1967). Physical connection approaches (Fair 1979; Salmon 1994; Dowe 2000)

try to analyze causation as involving the "transfer" of some quantity from cause to effect, with physics playing the role of providing a list of appropriate quantities (energy, momentum, etc.). Neither approach has much promise. Regularity approaches rely on an overly naïve conception of what laws are (for a corrective, see Maudlin (2004)), while physical connection approaches use tools far too blunt to capture causal facts at any level but the most microphysical, and probably don't even succeed there (see Hall forthcoming).

§4 Two varieties of nomological relations are worth studying

In the search for abstract nomological relations that will serve as the core of an account of causation, two guiding ideas are worth pursuing. According to the first, what distinguishes the causes of some event E is that they are the events upon which E's occurrence *depends*. According to the second, what distinguishes the causes is that they *suffice* for E, without redundancy.

The first idea is best developed by means of counterfactuals; the rival probabilistic approach (E depends on C iff, roughly, the probability that E occurs given that C does is greater than the probability that E occurs given that C doesn't) fairs poorly in a deterministic setting. In its counterfactual form, this dependency approach has dominated most of the best parts of the literature, beginning with Lewis's classic 1973 paper in which he first laid out his counterfactual analysis of causation. Here, the core relation was simple: E depends on C iff, had C not occurred, E would not have occurred. His recent 2004 attempts to fix up his original analysis by admitting a new variety of dependence: E depends on C (in this new way) iff, had the manner of C's occurrence differed in any of various ways, the manner of E's occurrence would have correspondingly differed. More interesting, in my view, are recent attempts by Yablo (2002), Hitchcock (2001), and others to articulate a relation of dependence that holds fixed certain factors in the environment of the candidate cause and effect: very roughly, their proposal is that C is a cause of E iff, holding fixed certain suitably chosen facts, E would not have happened if C had not happened. The hard work of these accounts goes into figuring out a general recipe for finding the facts to be held fixed, and providing precise semantics for this more complicated counterfactual construction.

Meanwhile, the sufficiency approach has lain almost entirely dormant since Mackie's (1965) bungling of it (though see Bennett 1988 for an attempted revival). I think it deserves much more sustained investigation. There's not much by way of contemporary literature to point to, here, so I'll try to get the ball rolling by sketching what I think is the right way to develop it.

First, the set S of causes of an event E should *collectively suffice* for that event, but should do so *non-redundantly*: i.e., no proper subset of S should

suffice for E. Then S had better not really include *all* the causes of E occurring at *any* time, since later ones will render earlier ones redundant, and vice versa. So amend, requiring that the set of causes of E that *occur at some given time* (before E occurs) non-redundantly suffice for E. What remains is to say what "suffice" means. Here is a first pass: A set S of events suffices for (later) event E just in case the occurrence of those events lawfully guarantees that E occurs. That won't do, since it will in general be possible for the events in S to occur jointly with some other "inhibiting" events that act so as to *prevent* the occurrence of E. A better idea is to say that S suffices for E just in case, were the events in S to occur *without any interference*, E would occur. If we agree that such interference would require the occurrence of some other, contemporaneous events, then we can simplify: A set S of events occurring at some time t suffices for (later) event E iff, were the events in S the *only* events occurring at t, E would (still) occur. Calling a set *minimally* sufficient just in case it is sufficient, but no proper subset is, we thus arrive at the following *updated sufficiency account*: C is a cause of E iff C belongs to a set of contemporaneous events that is minimally sufficient for E. That doesn't actually work (it's an easy exercise to show why). But it strikes me as a very good starting point.

§5 Where to look, and what to do

Enough stage-setting. I'll close with some suggestions for how best to get up to speed on the literature (not all of it, to be sure: but those parts that I think are most interesting and valuable), and a list of what strike me as some of the most urgent and interesting questions that need to be (further) addressed by philosophers of causation.

For deep background, Hume's *Enquiry Concerning Human Understanding* (especially sections 4–7) is a natural starting point. For more shallow background, the readings in Sosa and Tooley (1993) are quite valuable, although Lewis's "Causation" should be supplemented by the "Postscripts" that appear in his 1986. The literature following Lewis has been heavily driven by an astonishing array of examples, for which Hall and Paul's (forthcoming) *Causation and the Counterexamples: A Traveler's Guide* provides just that. Finally, the readings in Collins, Hall, and Paul (2004) survey the state of the art in that part of the causation literature (the best part, in my view) that sees causal facts as essentially facts about the counterfactual structure of the world.

A large number of fascinating questions about causation emerge from this literature. Here are just a few; collectively, they give a good sense of the fertility of this stretch of the philosophical landscape: What is the best way to develop a sufficiency approach? What is the best way to develop a dependency approach? What should we take the causal relata to be: events, facts, or both? What is the best way to solve the problems posed

by cases of redundant causation (where two or more processes separately guarantee the occurrence of some given event)? What is the best way to account for omission-involving causation (causation by omission, prevention, and causation by double-prevention, where an event C prevents something from happening, which had it happened would have prevented event E)? What is the best way to account for the asymmetry of causation? Is causation transitive? Is the causal structure of a process intrinsic to it? How many kinds of causation are there? How context-sensitive are attributions of causation? What sorts of laws allow for interesting causal relations? What proper foundations can be given to the theory of causal modeling?

This last question is worth pausing over. There has been, in recent years, a remarkable upsurge of interest in causation among psychologists, computer scientists, statisticians, political scientists, and other non-philosophers. A minor intellectual renaissance is brewing, and philosophers ought to occupy the forefront of it. Here is why: One important strand of research focuses on how statistical data can be used to draw inferences about causal structures. Central to this approach are "causal models", intended to represent systems of "variables" connected by "mechanisms". (These terms are all drawn from Pearl 2000, a very important recent book on the subject.) By careful appeal to and analysis of such causal models, it is possible to develop subtle ways of empirically testing causal hypotheses in light of statistical data. But a serious foundational problem as yet prevents this approach from attaining the kind of scientific rigor it ought to have. Crucial notions – most notably, the notion of a "mechanism" – are left almost wholly obscure, in a way which makes it impossible to say anything general or informative about what makes any given situation apt for description by one causal model rather than another. Earlier, I played up the importance of developing accounts of causation that have genuine theoretical utility. Here is a case in point: such an account is needed that can set this important area of research on a sound conceptual footing. Philosophers of causation, take up the charge.

References

Armstrong, D. M. 2004. "Going Through the Open Door Again: Counterfactual vs. Singularist Theories of Causation", in Collins, Hall, and Paul (eds) 2004. An earlier version of this paper appears in Sankey 1999.

Bennett, Jonathan 1988. *Events and Their Names*, Indianapolis: Hackett.

Collins, John; Hall, Ned; and Paul, L. A. (eds) 2004. *Causation and Counterfactuals*. Cambridge: MIT Press.

Davidson, Donald 1967. "Causal Relations", *Journal of Philosophy* 64: 691–703.

Dowe, Phil 2000. *Physical Causation*. New York: Cambridge University Press.

Fair, David 1979. "Causation and the Flow of Energy", *Erkenntnis* 14: 219–50.

Hall, Ned 2004. "Two Concepts of Causation", in Collins, Hall, and Paul (eds) 2004.

Hall, Ned forthcoming. "Causation: An Opinionated Overview", in F. Jackson and M. Smith (eds.), *Oxford Handbook of Contemporary Analytical Philosophy*.

Hitchcock, C. 2001. "The Intransitivity of Causation Revealed in Equations and Graphs", *Journal of Philosophy* 98: 273–299.
Hume, David 1748. *An Enquiry Concerning Human Understanding*.
Lewis, David 1973. "Causation", *Journal of Philosophy* 70: 556–67. Reprinted with postscripts in Lewis 1986a: 159–172.
Lewis, David 1983b. "New Work for a Theory of Universals", *Australasian Journal of Philosophy* 61: 343–377. Reprinted in Lewis 1999: 8–55.
Lewis, David 1986. *Philosophical Papers, Volume II*, Oxford: Oxford University Press.
Lewis, David 1999. *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press.
Lewis, David 2004. "Causation as Influence", in Collins, Hall, and Paul (eds.) 2004. An abbreviated version of this paper appears in *Journal of Philosophy* 97: 182–197.
Mackie, J. L. 1965. "Causes and Conditions", *American Philosophical Quarterly* 2: 245–264.
Maudlin, Tim 2004. "A Modest Proposal Concerning Laws, Counterfactuals and Explanations", unpublished ms.
Pearl, Judea 2000. *Causality: Models, Reasoning and Inference*. Cambridge: Cambridge University Press.
Salmon, Wesley 1994. "Causality Without Counterfactuals", *Philosophy of Science* 61: 297–312.
Sankey, Howard (ed.) 1999. *Causation and Laws of Nature*, Dordrecht: Kluwer.
Sosa, Ernest and Michael Tooley (eds.) 1993. *Causation*, Oxford: Oxford University Press.
Tooley, Michael 1990. "Causation: Reductionism versus Realism", *Philosophy and Phenomenological Research* 50, Supplement: 215–36.
Yablo, Stephen 2002. "De Facto Dependence", *Journal of Philosophy* 99: 130–148.