Causation: An Opinionated Overview

Ned Hall

§0 Introduction

Let's start with a sampling of philosophical accounts of causation. Famously, Hume (1748, section VII) took causation to consist in constant conjunction: an event of type E has an immediately preceding event of type C as its cause just in case C-events are always followed by E-events. Fast forwarding about 200 years, the core idea that causation consists in the instantiation of lawful regularities received an update from Davidson (1967), who argued that the regularity or law in question need not be explicit: the truth conditions for a claim of the form "event C caused event E" are simply that there be some feature of C and some feature of E such that the two events are "covered" by a causal law relating events with these features. Mackie (1965) took the regularity approach in a slightly different direction, arguing that C is a cause of E just in case it is an essential part of a condition lawfully sufficient in the circumstances for E. Lewis (1973) developed a very different idea suggested by Hume, drawing on successes in the clarification of the semantics of counterfactual conditionals in order to analyze causation by means of them: event E counterfactually depends on event C just in case if C had not occurred, E would not have; causation itself Lewis took to be the ancestral of counterfactual dependence. More recently (2000), Lewis modified this approach to incorporate a different kind of counterfactual relation. On the updated view, C can also cause E if, roughly, had the manner of C's occurrence differed in any of a large number of ways, the manner of E's occurrence would have correspondingly differed.

In the last 25 years or so, Lewis's approach has eclipsed the Mackie/Davidson approaches as a source of inspiration for work in the philosophy of causation: the literature houses dozens of variations on the counterfactual analysis of causation, whereas regularity accounts seem to have fallen into disuse. Meanwhile, a number of quite different approaches have taken root. Several authors (see for example Good 1961,1962; Suppes 1970; Salmon 1980; Eells 1991; Mellor 1995; Kvart 2004) have looked to *probabilistic* relations as the key to understanding causation, trying to build on the core key idea that causes should render their effects more probable. Strikingly different yet again is the approach of those who see causation as fundamentally a matter of a *physical connection* between cause and effect, typically involving the transfer of some quantity. Several writers (Fair 1979; Salmon 1994; Dowe 2000) look to physics for guidance as to what these quantities are: thus, Fair takes the crucial quantity to be energy; Dowe lets it be any quantity that, according to fundamental physics, is conserved. Ehring (1997) by contrast takes this approach in a more metaphysical direction, arguing that causation essentially involves the transfer of a trope, a particularized property.

However disparate, these approaches share a common reductionist commitment: for they are all at the very least compatible with the view that facts about what causes what reduce to some more basic set of facts—e.g., the totality of non-modal facts about the world, together perhaps with the facts about the fundamental laws that govern the world. Others (Tooley 1987, 1990; Armstrong 2004) argue instead that causal relations are themselves part of the fundamental ontology of the world, and facts about their instantiation do not even supervene on—let alone reduce to—more metaphysically basic facts. (Intimations of a view like this can be found in Anscombe 1971.)

That will do for now. As is evident from even so brisk a tour, philosophy of causation sprawls across a heterogeneous landscape of positions. The extent of the sprawl would be hard to convey even in a leisurely survey, given how deep run the differences in approach, aims, and methodology. One might see such pluralism as a sign of health. Unfortunately, a more accurate assessment warrants less optimism, for philosophers working within different paradigms tend to engage very little with one another. This state of affairs is all the more unfortunate, given what has been, in recent years, a remarkable upsurge of interest in causation among psychologists, computer scientists, statisticians, and other non-philosophers (see for example Pearl 2000). A minor intellectual renaissance is brewing, and philosophers ought to, but do not yet, occupy the forefront of it.

An overview like this is going to require a heavy investment in taxonomy; in providing one I will try to highlight those areas where philosophical disagreement about causation can most fruitfully be joined. Let's start, though, with a rare but important area of agreement.

§1 **Primary vs. secondary causal locutions**

English employs a wide variety of causal locations. Suzy throws a rock at a window, breaking it. Consequently,

(1.1) *Suzy* causes the window to break.

(1.2) Her *rock* causes the window to break.

(1.3) Her rock's *momentum* (as opposed to its color) causes the window to break.

(1.4) Her *throw* causes the window to break.

(1.5) The *fact that she throws* the rock causes the window to break.

(1.6) Billy is standing nearby; his *failure to stop her* causes the window to break.

Already, we have evidence that causes can be people, things, property instances, events, facts, and omissions. Similar examples suggest that effects can also be highly disparate—although, curiously, perhaps not *quite* as disparate as causes (try coming up with a natural

English sentence which describes a *thing* as an effect: it can be done, but only with some strain). Nor does the disparity end here, as witness examples like the following:

(1.7) Suzy's throw caused the window to break in part *because* the window was so fragile.

(1.8) Suzy's throw was *the* cause of the window's breaking.

There is a widespread consensus that the right way to handle such profusion is to single out a narrow class of locutions as the most important ones, the ones that an account absolutely has to get right. Dealing with the remainder is typically seen as a kind of mopping up action. This decision about which locutions to consider primary and which secondary typically goes hand in hand with some metaphysical claim to the effect that certain items are the *fundamental* causal relata; the cash value of this claim is that causal facts about other sorts of items are to be explained by means of, or somehow reduced to, causal facts involving the fundamental relata. Thus, I might take statements of the form "event C causes event E" to be primary, and announce that on my account events are the fundamental causal relata; that takes care of (1.4) automatically, and perhaps (1.5), if I can argue that it is merely elliptical for (1.4). I might then say that what makes *Suzy* a cause of the window's breaking is, roughly, that Suzy participates in an event—her throw—that causes the window's breaking; likewise for her rock. Sentences like (1.8) call for a different strategy, since it will hardly do to hold that the window's breaking has but *one* cause; I might therefore hold, as many do, that the definite article singles out not the *only* cause, but rather the cause most *salient* in the context.

§2 <u>The causal relata: events vs. facts</u>

I'm not done: (1.3), (1.6), and (1.7) still need treatment, and the question of which treatment is best leads straight into controversy. We will touch on that controversy in §6; for the moment, let us take on board the utility of this distinction between primary and secondary locutions, between more and less fundamental causal facts. Now comes our first major area of dispute: how should this distinction be drawn? The two serious candidates are these:

Events: The primary locution is "event C causes event E"; the fundamental causal relata are thus events.

Facts: The primary locution is "the fact that P causes it to be the case that Q"; the fundamental causal relata are thus facts.

In the literature, **Events** seems to be the odds on favorite. Regrettably, that literature fails to provide any particularly compelling reasons why this should be so. Often, one encounters the opinion that **Facts** does not even merit refutation, since, obviously, facts are abstract in a way that automatically renders them unsuitable to be causes or effects. But this opinion does not withstand scrutiny, as Bennett (1988) and especially McGrath (2002) have shown. Davidson (1967) offers a famous but fallacious argument that the logical form of causal claims militates against **Facts**; again, see Bennett for a rebuttal. What's more, the insistence that facts cannot be causes or effects has consequences whose awkwardness goes underappreciated. We'll review these issues in §6. Until then—and only to make the discussion less cumbersome—we will proceed under the assumption that **Events** is correct.

§3 Reductionism vs. non-reductionism

The second major source of disagreement will require a bit of setting up. Begin with the assumption that there is some fundamental ontological structure to the world, some set of truths about the world that serve as a basis to which all other truths about the world reduce (we need "reduce" here, and not the clearer "supervene"; for the latter notion does not build in the asymmetry of the former). Add that these fundamental facts divide into non-modal or categorical facts—e.g., about what objects there are, and what purely non-modal properties and relations they have at what times—together with facts about the laws of nature. Taking our inspiration from physics, let's go a bit further, construing the fundamental non-modal facts about our world to be the facts about what its complete physical state is at every moment, and construing "laws of nature" to be the fundamental physical laws that govern how these states evolve over time. (If we're going to be completely up-to-date, we should add to this inventory facts about the geometrical structure of spacetime, and remark that instantaneous physical states are to be understood as instantaneous states along some foliation of the spacetime.) Causal facts, then, will somehow reduce to these facts.

There are two broad classes of approach in the philosophy of causation that agree with this basic reductionist premise, and two broad classes of approach that dispute it. Let's take them in turn.

One way to be a reductionist is to look for what I call a "nomological entailment relation" that can hold between events, and then identify causation with that relation. The label is intended to be more evocative than tight-fitting, and rather than try to give some precise characterization I will begin with examples:

Crude sufficient condition account: C causes E iff C and E both occur, and from the fact that C occurs, together with the laws, it follows that E occurs.

Crude necessary condition account: C causes E iff C and E both occur, and from the fact that E occurs, together with the laws, it follows that C occurs.

Crude probabilistic account: C causes E iff C and E both occur, and the probability that E occurs, given that C occurs, is greater than the probability that E occurs, given that C does not occur.

Mackie-style regularity account: C causes E iff C and E both occur, and from the fact that C occurs, together with some suitable auxiliary premises describing contingent facts about the circumstances in which C occurs, together with the laws, it follows that E occurs; but this fact does *not* follow from the auxiliary premises and the laws alone.

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

Lewis-style counterfactual account: C causes E iff C and E both occur, and there is a (possibly empty) set of events {D_1, D_2,..., D_n} such that if C had not occurred, D_1 would not have occurred; and if D_1 had not occurred, D_2 would not have occurred;... and if D_n had not occurred, E would not have occurred.

In each case, the analysis displays something that we might, speaking loosely, call an *entailment* relation between the fact that the cause occurs and the fact that the effect occurs—where that relation is, crucially, mediated by the fundamental laws. (In the case of counterfactual analyses, this mediation is behind the scenes, and happens in virtue of the role that the fundamental laws play in fixing the truth-values of the counterfactuals; see for example Lewis 1973b.)

The physical connection views sketched earlier illustrate a second, different way to be a reductionist. The essential feature of these views is that causation consists (at least partly) in the transfer of something from cause to effect; the differences concern what that something is. Strikingly different from nomological entailment accounts though they are, such views clearly count as reductionist, at least on the plausible assumption that what gets transferred to what reduces to basic non-modal facts, together with the facts about what the fundamental laws are.

Anti-reductionism likewise comes in two flavors. One of these, for which I take as inspiration Shoemaker (1980) and Cartwright (1999), holds that the items that populate our world possess, as a matter of primitive, not-further-explicable metaphysical fact, certain causal powers: there is no interesting sense to be made of a totality of "non-modal" facts about the world to which causal facts (partly) reduce. What's more, on one very natural way of developing this view, laws are not some *additional* feature of reality, but rather serve as mere summaries—more less accurate—of the causal powers objects possess.

An example, to try to make the contrast clear: Both the reductionist and this kind of non-reductionist will agree that two negatively charged particles will exert a repulsive force on each other. (To keep things simple, pretend that these are Newtonian particles.) Here, crudely, is how the reductionist understands this claim: the two particles each possess a certain property, and there is, in addition, a fundamental law of nature which describes the total force on each particle partly as a function of its distance from the other particle and these properties. While it is certainly correct to say, on this view, that the particles possess a *causal power* to repel like-charged particles, that fact is, as it were, only *derivative* on the facts just mentioned. (Perhaps it reduces to a certain kind of counterfactual detailing what would happen, were the particles placed in proximity to each other, and were no other sources of force present.) The non-reductionist we are considering sees things differently. She holds that part of what it is for the particles to be negatively charged is for them to have the causal power to repel each other. In the order of being, as it were, this fact is primary.

To finish our survey, let us consider a very different way to be a non-reductionist. Here, Tooley (1990) and Armstrong (2004) are exemplars. Thus Armstrong, commenting on Lewis's work, writes

> In the course of these discussions Lewis introduces his justly celebrated "neuron diagrams". These involve imaginary systems of neurons that are hooked up according to certain often quite complex causal patterns of firings and inhibitions of firing. They are then used to illustrate such situations as epiphenomenal events, preemptive causation, causal redundancy and so on.
>
> When we look at these diagrams, we can immediately see that they are possible causal patterns, in most cases empirically possible patterns (you could construct such a circuit). Counterfactual theories of causation (and, it may be added, Regularity theories and Probability-raising theories of causation) struggle with these diagrams. Wittgenstein spoke of an open door that we had only to see and go through to escape philosophical confusion.... The solution that I recommend to the problems posed by the neuron diagrams is very simple. Where there is an arrow in a diagram showing that one neuron brings it about that another neuron fires, or is rendered incapable of firing, take it that here there is *a genuine two-term relation of singular causation* holding between cause and effect. Where there is no such arrow, deny that there is any such relation. This is the open door. (2004, italics added)

The surrounding context makes clear that Armstrong does not mean to be offering a theory of neuron-causation, but a theory of causation quite generally. Moreover, his talk of a "genuine two-term relation" is meant to signal that causal relations are part of the *fundamental ontology* of the world (on his preferred view, causation is a *universal*). There is no problem, on this view, with including in one's fundamental ontology an interesting, rich range of non-modal facts, nor with introducing fundamental laws to govern what non-modally happens. It is just that facts about what causes what need to be introduced as an *additional* ingredient of reality, not to be reduced—by counterfactual or any other means—to the remaining ingredients.

Let me now try to encapsulate the differences between the four types of philosophical account of causation. Imagine that you are a god, setting out to construct a possible world. According to the reductionist, all you need to do is to lay down what, non-modally speaking, happens, and lay down the fundamental laws that govern what happens; you will thereby automatically have generated all the facts about what causes what. Nomological entailment approaches say that you will have done so because you will have generated all the facts about the relevant sort of nomological entailment (counterfactual dependence, probability raising, or whatever). Physical connection approaches say that you will have done so because you will have laid down all the facts about which of the relevant quantities get transferred from where to where. By contrast, a non-reductionist of the second type thinks that your job is not done yet: you still need to specify which events instantiate the causal relation. And a non-reductionist of the first type thinks that what you have done has simply been misdescribed: there is no way for you to lay down a totality of facts about what, non-modally speaking, happens. You can introduce objects into the world and endow them with properties—but you *thereby* introduce causal powers into the world. Furthermore, having done so, there is no need for you to introduce laws; what laws there are is already automatically fixed by what sorts of things there are, and what sorts of properties—and therefore causal powers—they possess.

§4 <u>Methodological issues</u>

Can any considerations make one of these four approaches appear more compelling than the others? I think so. But to bring them out we will need to focus on two methodological questions: What is the proper *aim* of a philosophical account of causation? And what are the proper methods of investigation for achieving this aim?

The first question seems to have an obvious answer: the aim is to explain what causation *is*. Perhaps. But this answer sounds a lot better coming from a non-reductionist, especially one of the Tooley/Armstrong variety, than it does coming from a reductionist. The former sort of philosopher can offer up his account as part of a description and explication of the fundamental ontological structure of the world, and to be sure, no one with a taste for metaphysics would doubt the significance of such an enterprise. But matters are otherwise for the reductionist. By her own lights, her account of causation will add nothing whatsoever to a description or explication of the fundamental ontological structure of the world. That does not mean that her project is unimportant. It is just that its true significance will be masked by the glib claim that she is simply trying to explain what causation is.

Here is a comparison that might be helpful. Imagine that we are trying to figure out what life is—what the distinction is between living and non-living entities. Now, someone with vitalist sympathies will see this project as immensely important, for this is a distinction that, by his lights, carves nature at its joints. But from the standpoint of modern biology the significance of this enterprise lies elsewhere, and hinges entirely on the *utility* (for the discipline of *biology*, presumably) possessed by whatever distinction one comes up with. Similarly, I suggest that the most honest way for the reductionist about *causation* to advertise her project is as follows: in "analyzing" causation, she is trying to come up with a precise account of a distinction (or set of distinctions) that will serve some useful theoretical purposes (most obviously, in other areas of philosophical inquiry). The significance of this difference in methodological orientation between the reductionist and the non-reductionist will, I hope, become clearer as we proceed, and in particular, as we review the standard techniques that philosophers have used to support their views.

Turn now to the second question raised above, and consider the sorts of "data" that a philosopher has to rely on, when giving an account of causation. The data, not surprisingly, take the form of intuitions, which come in three varieties:

First, there are intuitions about the causal structures of specific hypothetical cases: one describes such a case, and then elicits intuitive judgments—ideally, firmly held, widely agreed upon intuitive judgments—about what causes what. Here is an example that will come in for a lot of discussion later on:

Suzy First: Suzy throws a rock at a window. It flies through the air, strikes the window, and breaks it. Suzy's friend Billy throws a rock at the window, too. He's slower, so her rock gets there first; but if she hadn't thrown it, the window would have shattered all the same, thanks to his throw.

The intuitive judgment that Suzy's throw is a cause of the window's breaking, and Billy's is not, fits the bill: it is extremely firmly held and widely shared. As we will see later on, it is also an very important intuition, for a number of attempts to give a reductionist account of causation founder on this case.

Second, there are intuitions about cases that are harder to classify, but that are called upon not so much to refute some specific proposal but to establish some sweeping claim about causation. Here I have in mind the cases that Tooley in particular relies on to motivate his variety of non-reductionism. A clear instance is the following (1990, italics in the original; I'll henceforth call this the "Tooley-case"):

> Given [the assumption that there is nothing incoherent in the idea of an uncaused event], one can consider a world where objects sometimes acquire property Q without there being any cause of their doing so, and similarly for property R, and where, in addition, the following two statements are true:
>
> (1) It is a law that, for any object x, x's having property P for a temporal interval of length Δt either causes x to acquire property Q, or else causes x to acquire property R;
>
> (2) It can never be the case, for any object x, that x's having property P for a temporal interval of length Δt causes x to acquire both property Q and property R.
>
> Suppose, finally, that an object a in such a world, having had property P for the appropriate interval, acquires both Q and R. In view of the law described in statement (1), either the acquisition of Q was caused by the possession of P for the relevant interval, or else the acquisition of R was so caused. But, given statement (2), it cannot be the case that the possession of P for the relevant interval caused both the acquisition of Q and the acquisition of R. So once again, it must be the case that one of two causal states of affairs obtains, but the totality of facts concerning, first, the non-causal

> properties of, and relations between, events, secondly, what laws there are, and thirdly, the direction of causation in all potential causal processes, does not suffice to fix which causal state of affairs obtains.

Grant the cogency of the case (a mistake, as I'll argue later), and Tooley's brand of non-reductionism follows at once.

Third, there are intuitions about the general principles that govern the causal relation. Here are some examples (to which both **Events** and **Facts** could be added):

Transitivity: Causation is transitive; if C is a cause of D and D of E, then C is thereby a cause of E.

Intrinsicness: The causal structure of a process is intrinsic to that process.

Locality: Causes are connected to their effects via temporally (stronger: spatiotemporally) continuous chains of intermediate causes.

Omissions: Absences—failures of events to occur—can be causes and effects.

Dependence: Counterfactual dependence suffices for causation: if distinct events C and E both occur, but E would not have if C had not, then C is thereby a cause of E.

Three comments about such principles are in order. First, it should not be thought that their intuitive appeal, such as it is, provides decisive grounds for endorsing them. This is partly because of the possibility of outright counterexamples. (For example, several cases in the literature purport to refute **Transitivity**.) But it is also partly because an insistence on them may carry severe theoretical costs: consider, e.g., the difficulty in wedding **Events** to **Omissions**. Second, these principles can receive independent support on the basis of theoretical considerations. For example, one reason for *endorsing* transitivity—even in the face of the alleged counterexamples—is that unless one does so one cannot construct an account that will get some clear and obvious cases right (See Lewis 1973a). Third, tensions may arise between various of these principles; for example, it is probably impossible to maintain both **Locality** and **Dependence** (see my 2002).

Obviously, one overarching aim for a philosophical account of causation is to accord as much as possible with the sorts of intuitions just canvassed. But there is bloody warfare about their relative priority. Tooley, unsurprisingly, leans great weight on the Tooley-case and others like it, whereas reductionists like myself are inclined to dismiss them out of hand. Again, Lewis takes it as a non-negotiable datum that facts cannot be causes or effects; Bennett sees this position as involving a crude mistake. Even when it comes to intuitions about the causal characteristics of hypothetical cases, opinions vary. Some philosophers, for example, see the purported counterexamples to transitivity as decisive, whereas others see them as merely misleading. (Compare, in this regard, McDermott 1995 with Lewis 2004.) Comparatively rare are example such as **Suzy First**, which, on the one hand, have a great deal of probative value, and, on the other hand, command intuitive verdicts so firm that virtually no philosophers would care to reverse them. So a philosopher of causation who wants to get beyond a foot-stamping defense of some position has her work cut out for her. In what follows, I will outline one plan for navigating the competing demands of the various intuitions just sketched.

§5 <u>The case for nomological entailment accounts</u>

The first step will be to argue the merits of a nomological entailment approach. Let's begin by reviewing three further constraints on an account of causation, independent of the constraint of "intuitive fit".

First, it would be desirable to have an account that applies across a wide range of possible worlds, especially worlds with different laws from our own. Causation is to a good extent "topic neutral”—at least when the topic is the fundamental laws of nature. We will see that this desideratum spells trouble for physical connection accounts.

Second, and more importantly, an account of causation needs to be “level neutral”: it needs to respect the insight that causal relations obtain at and across a wide variety of levels of description. I drink some coffee, and perk up as a result. There is causation at the level of my body: drinking the coffee causes me to perk up. There is causation at the biochemical level, involving the interaction of caffeine molecules with other molecules in my body. There is causation at the atomic and subatomic levels as well. And causation can cross levels: in suitable circumstances, the decay of a radioactive atom can bring about the birth of a child (by way of a romantic story involving two nuclear chemists, as it might be).

It is not enough merely to recognize the existence of causal relations at and across different levels. For these relations are not independent of one another: When I perk up as a result of drinking the coffee, for example, this body-level causal fact is not somehow completely metaphysically separate from and independent of the biochemical causal facts involving the behavior of the caffeine in my bloodstream. Now, the nature of this dependence is an intricate and subtle matter. All the same, an account of causation should not render it a complete mystery, and as we will see, a Tooley/Armstrong-style nonreductionism founders on this requirement. (We will also see trouble for physical connection accounts in even recognizing the existence of causal facts at any level other than the most microphysical.)

Finally, as noted above, we should hope for an account that does some useful theoretical work. Again, this desideratum is particularly important for reductionists.

Let’s turn to a more detailed look at the four broad classes of account, starting with Shoemaker/Cartwright-style nonreductionism.

For our purposes, the significant part of this position is the claim that there is no content to the notion of "non-modal fact"—or at any rate, not enough to meet the reductionist's needs. Why not? Because, so the story goes, to specify a thing's properties or relations to other things is—with metaphysical necessity—to at least partially specify its causal powers. (Even in the case of its spatiotemporal relations? Never mind.) Recall our electrons, exerting a repulsive force on each other in virtue of their negative charge. Could those electrons have existed in a world in which they possessed exactly the intrinsic physical properties that they in fact possess, but *attracted* each other? One might think not; one might think that to the extent that we have a grip on what it is for something to possess such-and-such a negative charge, it is inconceivable that two things could possess this property and not repel each other. That is, it is *essential* to this property that a thing possessing it have the causal power to repel other things possessing it. And once such a connection is recognized between fundamental physical properties and associated causal powers, it seems unlikely that this connection could be limited to the micro level. It is much more likely that it will percolate on up to the most macro of levels, so that even the possession of ordinary properties will be necessarily connected to the possession of certain causal powers. (Thus, Shoemaker contends that a knife's property of being sharp is necessarily connected to its power to cut).

But no harm will come to the reductionist from simply conceding the point. What she needs is some reasonably clear conception of the sorts of fundamental facts to which all other facts, and in particular causal facts, reduce. We have already provided such a conception: fundamental facts are facts about the geometry of spacetime, about the instantaneous physical states the world occupies, and about the fundamental laws that govern their evolution. She might wish to claim, in addition, that this package of ingredients has a purely non-modal component (instantaneous states, plus spacetime geometry), and a modal component (the fundamental laws). The Shoemaker/Cartwright position, if correct, shows that she can't, by showing in essence that there is no way to specify a complete instantaneous physical state of the world without placing substantial constraints on the fundamental laws. So what? The additional claim is an optional extra, and abandoning it leaves untouched what is central to the reductionist picture.

That is an important result. Certainly, the Shoemaker/Cartwright view is *compatible* with a kind of primitivism about causation that stands in direct opposition to reductionism—namely, the kind that holds that all that can be said about causation is that objects in the world possess various causal powers that are manifested on various occasions, with no prospect for an informative general theory of how these powers are constituted, and likewise no prospect for an informative general theory of the causal relation. This view strikes me as pointlessly defeatist: at best it is a position to end up with, after all other avenues have been thoroughly and exhaustively investigated. It is crucial to see, then, that it gains no support from what may well be an attractive position about the metaphysics of properties.

Before proceeding to the second, Tooley/Armstrong variety of nonreductionism, it will be helpful to review a well-known and important distinction between two conceptions of properties and relations. On the first, "non-discriminating" conception, properties are cheap: It is not merely that for every (suitably non-pathological) predicate, there is a property or relation that is its extension, it is that for every function mapping possible worlds to sets of n-tuples of objects that exist in them, there is an n-place relation. Roughly, any way of classifying yields up a property or relation, and thus there is no sense in which properties and relations help carve nature at its joints.

On the second, "sparse" conception, what counts as a property or relation is vastly more restricted. In my view, the guiding intuition for this conception is the following: what properties and relations there are at a world is limited by the requirement that there be just those properties and relations that are needed to exactly specify what is true at that world, and no more. (See Lewis 1983 for additional characterizations.) Thus, if there is the property of being F, and the property of being G, then there is no such property as the property of being either F or G; for the pattern of instantiation of F-ness and G-ness will automatically fix the pattern of instantiation of this other "property", and so its addition to the list of properties would be redundant in a way that disqualifies it.

Two points, by way of clarification. First, the guiding intuition is only that, since the foregoing requirement can obviously be met by distinct families of properties and relations. Second, a combination of physicalism with a modest brand of scientific realism yields the view that one of the aims of science—fundamental physics in particular—is to provide us with an inventory of the properties and relations, sparsely conceived, that are instantiated in our world.

The Tooley/Armstrong view is a thesis about properties and relations, sparsely conceived. It is the thesis that the list of such properties and relations includes causation as one of its members. That is clearly their intent, and at any rate to understand their position as having in mind the non-sparse conception would render it utterly trivial, since everyone agrees that on that conception, causation is one of the relations that exist. (Even those who think that there is no such thing as causation: for they will simply take it to be the empty relation.)

One significant motivation for this view comes from a certain kind of thought-experiment, of which the Tooley-case is an instance. If successful, these thought-experiments would appear to establish that facts about what causes what do not even supervene on, let alone reduce to, facts about the fundamental laws, together with the facts about the pattern of instantiation of fundamental properties and relations. The most natural response? Insist that causation simply needs to be included on this list of fundamental properties and relations.

But the value of these thought-experiments is doubtful. First, the cases tend to be described in ways that are question-begging, at least if the opponent is the reductionist. Notice that it was crucial to Tooley's example that the statement of the laws explicitly involved the notion of causation. To see this, try running through the example with "causes" replaced by "is followed by". Do that, and the case can no longer be described consistently, for we are told that it is a law that being F is followed by exactly one of being G and being H, and also that it is compatible with this law that being F can be followed by both being G and being H. So, specifying the content of the law in the way that he did was crucial for the success of Tooley's example.

But a sober-minded reductionist should not accept that the fundamental laws to which she thinks the causal facts in part reduce are ever specified in such terms. To be sure, laws of the special sciences might need to be specified in part by appeal to the notion of causation; for example, it is a law of metallurgy that, in absence of interfering factors, heating a metal bar causes it to expand. But the fundamental laws are different. Our best models for them, after all, come from fundamental physics, where these laws are typically captured in the form of differential equations, or in the form of recipes for constructing differential equations (e.g., Hamilton's equations in classical mechanics, or the Schrödinger equation in non-relativistic quantum mechanics). And the content of these laws is exactly that such-and-such complete physical states are followed (with nomological necessity, if you like) by such-and-such other complete physical states.

Second, even if Tooley succeeds in showing that our ordinary concept of causation carries with it nonreductionist commitments, the reductionist is well within her rights simply to respond, "so much the worse for our ordinary concept". Compare the case of "life": you can tell an understandable, conceptually coherent story according to which a rock is alive, even though it has the internal physical structure of, well, a rock. But no concept of life as articulated by biologists or philosophers of biology will allow that this is a genuine metaphysical possibility. It is laughable to suggest that this points up some deep flaw in the work of those biologists and philosophers of biology. Rather, what it shows is that to the extent that we want a precisely articulated concept of life that will serve some useful purpose, the structure of our ordinary intuitions about the word "life" are at best an imperfect guide to the structure that concept should have. A reductionist can and should say the same thing about the concept of causation.

So much for one of the main motivations for the Tooley/Armstrong view. What other motivations are there? As far as I can see, the only important remaining one consists in despair at the prospects for any other, more illuminating account of causation (cf. the passage from Armstrong quoted earlier). In response, I do not think that the full range of resources available to reductionists have yet been appreciated, let alone exhausted; §6 will indicate why.

So much for the reasons in favor of nonreductionism. On the other side of the balance sheet lie two very serious problems. The first concerns the requirement of level-neutrality—the requirement that an account of causation should respect the existence of causal relations at all levels from the most microphysical to the most macrophysical. Now, it is not that the nonreductionist cannot respect the existence of such relations; for example, when I drink the coffee there is, on his view, a two-place singular relation of causation that obtains between the event of my drinking and the event of my perking up, and likewise a multitude of instances of this relation at the biochemical level (between events involving caffeine and the cells of my body). What is a complete mystery is why there should be any relations whatsoever of metaphysical dependence between the patterns of instantiation of causation at these different levels. Minimally, it seems that we should have this much: if in any world the pattern of events in my body is exactly as it actually is, and the pattern of instantiation of causal relations at the biochemical level is exactly as it actually is, then so too the events of my drinking the coffee and my perking up instantiate the causal relation. But if all we say about the causal relation is that it is one of the fundamental relations instantiated in our world, then our view provides no resources for explaining why this relation of metaphysical dependence should hold.

Compare a simple counterfactual analysis. It is rather plausible that from the claim that the events in my body are exactly as they actually are, and the claim that the patterns of counterfactual dependence at the biochemical level are exactly as they actually are, it will follow (bracketing pre-emption: see §6) that my perking up counterfactually depends on my drinking the coffee. By contrast, the Tooley/Armstrong view leaves the connections between causation at different levels unacceptably mysterious.

The second problem is epistemological. What, on the nonreductionist account, could possibly be the evidence for any causal claim? This problem famously dates back to Hume, who argued that neither perception nor introspection acquaints us with any instance of a “necessary connection”. Now, one might doubt Hume’s sweeping claim, and think for example that in the case of the impacts of various objects on one’s body one directly perceives causal relations. But that possibility is a red herring, as witness the fact that in a wide variety and perhaps the vast majority of cases, our knowledge of or justified beliefs about causal relations are inferential. You flip a switch, and a light goes on. You flip the switch back to its original position, and the light goes off. You repeat the experiment a hundred times, and the same correlation obtains. You infer that flipping the switch (say, the first time) *caused* the light to go on. You did not perceive this causal relation. All the same, your inference certainly seems reliable. But it is a mystery why, on the nonreductionist view, it should be.

Isn't such underdetermination of theory by data everyone's problem? No: not underdetermination of *this* variety. Compare garden-variety underdetermination of the kind that yields to a familiar application of inference to the best explanation: we believe that there are atoms, for example, because of the stunning success of a worked-out atomic hypothesis at explaining a vast range of observable phenomena. Whatever you think of this style of inference, you can at least agree that we do not have a *best* explanation unless we have *an* explanation. Well, what observable phenomena does the nonreductionist *explain* by means of the hypothesis that events (e.g., the flipping of the switch and the light's going on) instantiate some fundamental, two-term relation? None, as far as I can see. Calling this relation "causation" only helps hide this failure from view.

Let's turn now to physical connection accounts. A common motivation for them is that they allegedly make the investigation of causation properly empirical (see for example Fair 1979). The thought, apparently, is that "cause" is a natural kind term, and what is needed in order to give an account of causation is some empirical hypothesis about what relation causation is. But this claim seems to involve a basic confusion about the semantics of causal terms. Here I am content to follow Schaffer's excellent discussion (2004). He points out that our semantic intuitions about hypothetical cases involving laws quite different from the actual laws show that "cause" and other related terms do not function in anything like the same way that terms like "water", " gold", etc. do. This observation reveals, in addition, an important limitation on physical connection accounts, which is their failure to generalize across a wide range of other possible laws. It is not merely that specific such accounts (e.g., that causation involves transfer of momentum) are hostage to physics as we currently conceive it, but that causal relations can, apparently, obtain even in worlds whose laws do not pick out any quantity as one that might plausibly be "transferred" in all and only causal interactions.

It is sometimes thought that physical connection accounts gain support from pre-emption cases like **Suzy First**. For, allegedly, the only thing that distinguishes Suzy's throw from Billy's is that there is a momentum-carrying process connecting her throw to the shattering which, at the moment of the shattering, results in a transfer of momentum from the rock to the window—whereas there is no such process connecting Billy's throw to the shattering. (Ehring leans heavily on cases such as this to motivate his account.) But this motivation is misguided, and seeing why points out another serious problem with physical connection accounts. Add some detail to the case: Billy's rock, as it moves through the air, pushes air molecules ahead of it, and those bump into other air molecules, etc. And let us suppose that some of the air molecules thus moved strike the window at the exact moment that Suzy's rock strikes it, transferring some small quantity of momentum to the window. According to physical connection accounts, the revelation of these extra details ought to reverse the original judgment, turning this into a case of symmetric overdetermination: *both* Suzy's and Billy's throws count as causes of the window's breaking, since both are connected to it by the transfer of some appropriate quantity. But that is plainly false. Transfer of some quantity is thus not enough for causation; the account needs to be supplemented by some story about what counts as a *relevant* transfer of the quantity. What's more, in a world as chock full of stuff—e.g., photons and other tiny, fast-moving particles—as ours, to say that transfer of some quantity is *required* for causation is to say virtually nothing at all, for, plausibly, every event will be connected in such a way to every other event in its future light cone. It thus seems that all of the work will be done by the account of "relevant transfer".

On the face of it, the most plausible way in which to develop such an account is by appeal to counterfactuals or perhaps probabilities—exactly the tools that a nomological entailment account will wheel out.

The problem just raised arises even at the most microphysical level: imagine a situation, for instance, in which particles A and B simultaneously collide with particles C, and particle C reacts in some way—but given the distinct physical characteristics of A and B, it is only the collision with A that *causes* this reaction. But even if this problem is somehow solved at the microphysical level, it won't go away at the macrophysical level. To make this clear, let's pretend that the microphysics is as congenial as possible to a physical connection account: all that happens at the microphysical level is that tiny, perfectly rigid billiard balls collide with each other in perfectly elastic collisions. The physical connectionist can thus say that at the microphysical level, causation consists in the carrying or transfer of momentum. Meanwhile, here at the macrophysical level I am alternately sipping coffee and orange juice, and a bit later perk up. It was sipping *coffee* that caused me to perk up, not sipping orange juice (my blood sugar was fine, I just needed a caffeine dose). But there is, patently, no way to account for this causal difference in terms of transfer of momentum: both the events result, by complicated routes, in the transfer of momentum to the particles that make up my brain, for example. Nor, to further dramatize the point, would one be able to account for the fact that is the *caffeine* in the coffee that perks me up, not (for example) the coffee temperature—even though, in virtue of the temperature, a bit more momentum gets transferred into the particles of my body than would have been the case if the coffee had been cool. The upshot of this sort of example should be obvious: even if a physical connection account works at the microphysical level, there is no plausible way to extend it into an account of causation at any other level. The requirement of level neutrality is thus violated in a rather stark manner.

Time to consider one final view, before turning triumphantly to nomological entailment accounts. This view, for reasons that will emerge shortly, is hard to characterize with respect to the reductionism/nonreductionism distinction. The following passages from Menzies will serve as an introduction (1996, pp. 97 and 101; italics in the original):

> The first step in providing a … definition of the causal relation is to set down the central tenets of our folk theory of causation—the platitudes about causation which are common knowledge among us. There are many such platitudes: for example, it typically coincides with a temporal ordering of events so that causes precede effects; it typically coincides with the means-end relation so that if an effect is an end, its causes are means to it; causes explain their effects. The *postulate of the folk theory of causation* will consist of a conjunction of all such platitudes; or better, a long disjunction of all conjunctions of most of the platitudes. The most important platitudes—the ones which are crucial to the concept of causation—will be elements in all the conjunctions in this long disjunction. In the subsequent discussion I concentrate on a simple formulation of the postulate of the folk theory of causation, a formulation which takes the postulate to be a conjunction of three crucial platitudes.
>
> …
>
> These three crucial platitudes, then, will be taken to form the basis of the postulate of the folk theory of causation. On the basis of this postulate, we can frame an explicit definition of the causal relation: The definition runs: the *causal relation is the intrinsic relation that typically holds between two distinct events when one increases the chance of the other event*. (The notion of increase in chance, used here, is to be understood in the counterfactual manner elucidated in Lewis's theory….)

(The platitudes are that causation is a two-place relation between distinct events, that it is intrinsic to its relata, and that it is connected in the way indicated to counterfactual chance-raising.)

Consider the abstract form of the proposal (drawn, as Menzies makes clear, from the treatment of theoretical terms in Lewis 1970): We focus on some domain of relations. We exhibit a set of characteristics, drawn from the “folk theory of causation”, that relations in the domain might or might not possess. We identify causation as the unique relation in the domain possessing these characteristics. Why proceed in this way?

Menzies’s principal reason is that he is impressed by what looks to be the utter failure of reductionist accounts to handle cases such as **Suzy First**, and probabilistic variants thereof. The strategy of taking “causation” to be a theoretical term, amenable to the kind of treatment just sketched, seems a promising alternative: what distinguishes Suzy’s throw from Billy’s as a cause of the window’s breaking is that the pair <Suzy’s throw, breaking> instantiates the unique relation which plays the role picked out by our folk theory of causation, whereas the pair <Billy’s throw, breaking> does not.

But there is less here than meets the eye. First, a minor problem: in all likelihood *nothing* will fit Menzies’s definition, because of the unfortunate requirement that causation be intrinsic to its relata. This fuse is lit; a bit later, that bomb explodes. Suppose the first event causes the second. That is does so is surely partly determined by features of the situation *extrinsic* to the lighting and the explosion—e.g., that this fuse is *connected to* that bomb.

Set this problem aside—perhaps by removing or suitably amending the requirement of intrinsicness. A more serious problem remains. Zero in on the first step: what, exactly, is the domain of relations over which we should quantify?

Perhaps it is the domain of fundamental relations, that is, relations understood in accordance with the *sparse* conception of properties and relations. But then the proposal turns into Tooley/Armstrong-style nonreductionism, with extra bells and whistles that add nothing of value.

Perhaps, then, it is the domain of relations, understood in accordance with the *non-*

discriminating conception of properties and relations. But then a dilemma arises. On the one hand, the definition offered might be so demanding that no relation can fit it. On the other hand, if any relation fits it, then the account will become trivial, since the definition will pick out this relation *too directly*. Let me explain, by way of a simple illustration.

Suppose Menzies drops the intrinsicness requirement, so that the causal relation is, now, defined to be the relation that typically holds between two distinct events when one increases the chance of the other event. Now it is indeed possible for a relation to fit this definition. In fact, here is one relation—a relation in, remember, the non-discriminating sense—that fits it as well as could be desired: it is the relation that holds between two distinct events when one increases the chance of the other event. But *this* relation cannot be what Menzies has in mind: for if it is, then we have made no progress on the preemption problem that originally motivated his approach. Making the definition more sophisticated won't help in the slightest: for again, if it can be met by any relation, it can be met by that relation that holds between distinct events exactly when …, with the blanks filled in by the definition itself. In short, we can see that the use of the word "typically", and indeed of the whole apparatus of theoretical definition, is doing *no work*: if the account succeeds, it succeeds because the definition in fact *constitutes a successful analysis of causation* of the kind Menzies thinks is unattainable.

Might there be a middle way? That is, might there be a way to specify boundaries on the relations being quantified over that are broad enough to distinguish Menzies's approach from the Tooley/Armstrong view, but narrow enough to avoid the foregoing trivialization? Perhaps. Menzies, for example, tries to appeal to a criterion of "naturalness". But what is unnatural about the relation of chance-raising? Again, one might restrict the domain to relations that can be described, simply, using the resources of fundamental physics. But then the account runs afoul of the problems that beset physical connection accounts. My own suspicion is that the project is misconceived: the idea that we get any illumination into the nature of causation by conjoining the central platitudes involving the concept, and defining causation as that relation which satisfies them, is misguided. Of course we will gain illumination by *figuring out* what the central platitudes are. But we gain nothing further by trying to leverage them in the way Menzies does into a definition of causation. Better, I think, to return to the honest toil of trying to come up with an explicit, reductive analysis of this notion.

The next sections will consider the most promising options for such an analysis, and some of the most striking obstacles they still need to overcome. But let us first review some of their principal strengths. I will do so by considering, purely by way of illustration, the simplest counterfactual account, according to which event C is a cause of distinct event E just in case, if C had not occurred, E would not have occurred (the lessons generalize straightforwardly to more sophisticated accounts).

The account is intuitively attractive, not only because it works for a very wide range of cases, but also because its seems to get those cases right for the right reasons. What's more, it cleanly meets our desideratum that a theory of causation should be neutral as to the precise content of the laws: all the account requires of the laws is that, together with the actual facts, they fix truth values for the needed counterfactuals, and standard semantics for counterfactuals (cf. Lewis 1973b) show how this can be done provided only that the laws draw a distinction between those worlds that are nomologically possible and those that are not. Finally, level-neutrality poses no problem whatsoever, since it is routine that macrophysical events such as my perking up can counterfactually depend on other macrophysical events such as my drinking the coffee. And, as we've already seen, there is no deep mystery why there should be interesting relations of metaphysical dependence between causal facts at different levels. None of this is to say that the counterfactual account is correct, for it's not. It is rather that before one complains too bitterly about the problems that beset it and other reductionist nomological entailment accounts, one should keep vividly in mind how deep and systemic are the problems that beset the major rivals.

There is, finally, a very different but important reason for thinking that the philosopher interested in understanding causation would best spend her efforts by investigating the prospects for some nomological entailment account. And that is because even if she fails to come up with such an account, the lessons learned along the way—about such topics as varied as the nature of the causal relata, the varieties of causal pre-emption, tensions between basic theses involving causation such as **Transitivity**, **Intrinsicness**, and **Dependence**, the relation between ordinary causation and causation involving omissions, and much more—are rich enough to repay these efforts many times over. To put the point another way, some failures are more interesting than other failures; the problem with the rivals to nomological entailment accounts is not merely that they fail so disastrously, but that they do so in a boring fashion. One learns little of value from their demise. What follows, though regrettably brief, will try to point out some of the most important ways in which matters are otherwise, for nomological entailment accounts.

§6 Nomological entailment accounts and the crucial examples: preemption, omissions, and transitivity

First we need to drastically narrow the scope of discussion. As anyone with even a passing acquaintance with the literature in this area knows, it is heavily driven by examples, and by now they are legion. We will consider just a handful of those examples that have the most to teach about the proper form of a theory of causation: certain kinds of preemption, certain examples involving omissions, and certain purported counterexamples to **Transitivity**. (For a much more thorough survey, see Hall and Paul forthcoming.)

A theory of causation needs a guiding idea, and two seem most worth taking seriously. The first is that the occurrence of the cause *makes a difference to* the occurrence of the effect. The second is that what distinguishes the causes of some effect is that they collectively *suffice* for that effect. Each of these ideas can be developed in a number of different ways, but to fix ideas—and, again, to keep the discussion manageable—let us assume that the development of the first identifies causation with some sort of counterfactual dependence: given some event E, its causes at some earlier time t are to be those t-events on which it depends (leaving open whether this is simple counterfactual dependence, or some more complicated variety). As to the second, we should understand it as building in some requirement of nonredundancy, so that what distinguishes a set of causes is not merely that the elements collectively suffice for the given effect, but that no proper subset of them does. So: given some event E, its causes at some earlier time t are those t-events collectively minimally sufficient for t (leaving open exactly how "sufficient" is to be defined). Let us henceforth label these approaches the *dependency* approach and the *sufficiency* approach.

Are causes and effects here *events* or *facts*? We should leave that for theory to decide, resisting the widespread assumption that the "abstractness" of facts offends against taking them to be causal relata. For this bias is utterly without justification. Observe, first, that as soon as you attempt to construct a nomological entailment account of causation, facts are already implicated, even if only facts of the form "the fact that event C occurred". After all, when we talk, for example, of counterfactual dependence of event E on event C, that is really derivative: in the first instance, counterfactual dependence is a relation between the fact that event C occurred and the fact that event E occurred. Second, the automatic removal of this option from the table makes for trouble when we consider causation involving omissions.

Granted, matters are otherwise for Tooley/Armstrong nonreductionism, or for physical connectionist reductionism. Nonreductionism treats causal relations as part of the fundamental ontology of the world; plausibly, relations of this sort need certain sorts of relata. For example, if one is convinced (with Armstrong) that relations of this sort are universals, then they need particulars as relata, and we might grant that facts are not particulars. Again, if one thinks of causation as involving the transfer of something from cause to effect, then cause and effect need to be the sorts of things that can transfer and receive stuff. Perhaps events can do so, but facts can't. However, once those two approaches have been abandoned, and we have turned instead to approaches that see causation as, at heart, involving some relatively abstract relation of some sort of nomological entailment, then the ontological scruples that the two abandoned positions motivate should likewise be abandoned. Continued concern over the "abstractness" of facts makes no sense.

Concern of another sort *does* make sense, at least if one adopts the sufficiency approach. For unless we restrict which sorts of facts are allowed to be causes and effects, the existence of purely logical relations among facts will make it impossible to discern when facts stand in the proper relations of sufficiency. Kim (1971) makes this point very effectively against Mackie, and it is worth reviewing here. Consider a case where Suzy throws a rock at a window, breaking it, and at the same time Billy throws a rock at a different window, breaking it as well. We want it to come out that the fact that Billy threw his rock causes it to be the case that his window broke. Let P be the fact that he threw his rock, and Q the fact that his window broke; then we need to find some set which includes P and is minimally sufficient for Q. Suppose we have done so. Then, regardless of how we analyze "sufficient", it would seem that we can construct another set, as follows: let R be the fact that Suzy threw her rock, and simply replace P by the pair (R, either not-R or P). The resulting set will likewise be minimally sufficient for Q: so the fact that Suzy threw her rock causes it to be the case that Billy's window broke. That is a disaster, and one way to circumvent it is to insist that facts can count as causal relata only when tied to particular events—that is, they must be facts of the form, "the fact that event C occurred". The fan of **Events** would be well within her rights to complain that there is little difference between this position and the position that it is events that are the fundamental causal relata.

Still, this is only one move in a debate that really needs to be allowed to run its course. First, it is curiously difficult to advance a similar argument against the combination of **Facts** with a dependency approach. Second, similar problems will plague the fan of **Events**, unless she is equipped with a theory of events that will enable her to distinguish causal from non-causal relations between them (cf. Kim 1973; for a response, see Lewis 1986c). Finally, troubles of a different sort arise for the fan of **Events** when she insists on forcing certain sorts of causal locutions involving omissions into her preferred mold.

Of this more shortly. For ease of exposition, we will pretend, until we can no longer do so, that causes and effects are events. Time now to turn to the examples.

Begin with cases of redundant causation, which feature the presence of backup processes poised to bring about the given effect. Such cases come in a number of distinct flavors (see Hall and Paul 2003 for a comprehensive survey). Here I will consider just one variety, sometimes called "late preemption", and exhibited in **Suzy First**. One causal process brings about some event E (in this case, the window's breaking), and by doing so, prevents the backup process (initiated by Billy's throw) from itself bringing about this event. (Contrast cases of "early" preemption, where some stage of the main process itself cuts short the backup process. For example, it might be that Billy, when he sees Suzy winding up to throw, decides not to throw himself.)

Cases of late pre-emption are utterly mundane, evoke rock-solid intuitions (Suzy's throw is a cause of the window's breaking; Billy's is not), and are extraordinarily difficult to handle, for the simple reason that for a very wide range of nomological entailment relations, the genuine cause and the idle backup will be symmetric with respect to that relation. Thus, both Suzy's throw and Billy's throw are elements of sets of events minimally sufficient for the window's breaking, the window's breaking counterfactually depends on neither, etc.

Rather than go through an exhaustive list of the various attempts that have been made to break this symmetry, let me simply fast forward to what I think are by far the two most plausible approaches. The first is a dependency approach, recently developed by Yablo (2002) and Hitchcock (2001), as well as others. It aims to show that the effect *does* differentially depend on the cause, provided one is careful to specify the right form of dependence (which, following Yablo, we will call "de facto dependence"). The guiding idea is that, *holding fixed* how certain aspects of the situation actually play out, the effect depends upon the cause, but does not depend upon the idle backup. In **Suzy First**, for example, the counterfactual that bears witness to the causal status of Suzy's throw is something like the following: if Suzy had not thrown, and if (as actually happens) Billy's rock had still not struck the window, then the window would not have broken. The idea is that there is no similar way to construct a counterfactual that will bear witness to the causal status of Billy's throw—since, as it actually happens, Suzy's rock *does* strike the window. I think this idea is attractive enough to deserve serious exploration, although it has yet to be developed in a fully adequate form. Part of the reason is that we need a rigorous recipe for picking out what, in the actual situation, should be held fixed as one moves to the counterfactual situation in which the candidate cause C does not occur. (Both Yablo and Hitchcock go some way toward providing such a recipe, but more work needs to be done.) Second, we need clear truth-conditions for the sorts of counterfactual used in this account, since it is not at all obvious how to evaluate what goes on in a world in which Suzy does not throw, Billy does throw (for that fact is, all agree, to be held fixed), but Billy's rocks somehow fails to strike the window.

The second strategy for handling cases of late pre-emption tries to build on the thesis that, roughly, the causal structure of a process is intrinsic to it. The idea, in outline, and illustrated by way of example, is as follows: Consider a case where Suzy is all by herself, and throws a rock at the window, breaking it. One could reasonably expect that an account could get the causal structure of *that* situation right. But in **Suzy First** there is a structure of events connecting her throw to the window's breaking that is *intrinsically just like* the process that unfolds in this simpler case. So if an analysis can get the causal structure of the such simpler cases right, then perhaps it can discern the causal structure of more complicated cases by finding in them intrinsic matches of what goes on in the simpler cases. But here too there is work to be done. For example, in **Suzy First** there is also a structure of events connecting Billy's throw to the shattering that is intrinsically at least somewhat like what goes on in a situation where Billy alone throws. So the strategy, if it is going to work, needs to develop a notion of intrinsic similarity *in relevant respects*. (See Hall 2003 for an attempt to do so.)

Let us turn now to cases involving omissions. These come in at least three varieties; the following case illustrates all of them:

The trip: Joey, seeing Suzy about to throw a rock at a window, runs toward her to try to stop her. Unfortunately, he trips over a tree root, and consequently doesn't reach her in time. She throws. The window breaks. If he hadn't tripped, he would have stopped Suzy, and the window would not have broken.

Here we have causation *by* omission: Joey's failure to stop Suzy is a cause of the window's breaking. And causation *of* omission: the trip prevents Joey from stopping Suzy (i.e., causes his failure to do so). And, finally, causation *by way of* omission (aka causation by *double-prevention*): Joey's trip is a cause of the window's breaking by way of his failure to stop her (his trip prevented him from stopping her; had he done so, the window would not have broken).

A clean treatment of these varieties of causation that merely extends some treatment that works for more ordinary cases has proved enormously difficult to come by. There are a number of reasons.

First, our judgments involving causation by omission in particular seem peculiarly sensitive to *normative* considerations. Example (from Sarah McGrath): Suzy goes away on vacation, leaving her favorite plant in the hands of Billy, who has promised to water it. Billy fails to do so. The plant dies—but would not have, had Billy watered it. So Billy's failure to water the plant caused its death. But Vladimir Putin also failed to water Suzy's plant. And, had he done so, it would not have died. So why do we not *also* count his omission as a cause of the plant's death? Because, unlike Billy, he made no promise to water it. But what does such a normative consideration have to do with *causation*? A natural response is: nothing—the causal structure of the example is what it is quite independently of who promised what to whom, or indeed of any other normative consideration. But then our ordinary judgments involving causation by omission commit deep and systematic errors. The literature, when it notices this problem, tends to badly underestimate its difficulty. (See Beebee (2003), and especially McGrath (2002) for notable exceptions.)

Second, an approach that insists that causation relates *events* will, confronted with omission-involving causation, lead to a badly distorted ontology. That is because omissions are not, in any decent sense, a species of event, and treating them as such yields only confusion. Ordinary events can be located in space and time, involve particular objects as constituents, have other ordinary events as parts, possess detailed intrinsic natures. None of this holds for omissions. Granted, you might think that x's failure to F is located wherever x is, at the relevant time (e.g., Billy's failure to water Suzy's plants takes place wherever he is, at the time that he should have been watering the plants). But there may be no "relevant" time. (McGrath again: Billy never runs a triathlon. When does *that* happen?) And there may be no "x". Billy, let us imagine, dies in a tragic accident, and so fails to meet Suzy for lunch the next day. Where does this omission take place (granting, for the sake of argument, that it takes place during lunchtime)? Where Billy is? But he isn't. You could, perhaps, make up an answer. But if so, you ought in good conscience recognize that that is all you are doing.

Better to treat causation involving omissions as causation involving a certain kind of *fact*—namely, a fact to the effect that no event of such-and-such type occurs. This approach gains confirmation from the way we support claims of causation by omission. We typically do so by way of an appeal to a counterfactual of the form, "if an event of type C had happened, event E would not have happened." From such a claim one can straightaway infer (modulo the worries about normative considerations) that the failure of an event of type C to occur caused E to occur. What we *don't* do is to pick out some particular event C, and say that if only *it* had occurred, then E would not have. The friend of **Facts** has a distinct edge, here.

More interesting results begin to emerge when we consider how a sufficiency-approach might be extended to cover causation involving omissions. The problems seem dire, and come quite clearly into view if we adopt a natural analysis of "sufficient": say that a set of time t events S is sufficient for later event E iff, had only the events in S occurred at t, E would (still) have occurred. If so, can a set of events be minimally sufficient for some omission? No: for the *empty* set will itself be sufficient. Can a set of events minimally sufficient for E *include* an omission? No: for even if we treat it as an event, its removal from S will not, on the given analysis, affect the set's sufficiency for E. Finally, consider a set S sufficient for E that includes a "double-preventer": an event C that prevents something that would in turn have prevented E. Then removing C will not affect sufficiency, since the test for sufficiency asks us to consider a situation in which the prevented threat to E is *also* absent. Should we blame the analysis of sufficiency? I don't think so: other attempts to flesh out this notion either remain hopelessly vague (cf. Mackie 1965) or run into worse problems (for discussion, see Hall forthcoming).

Things look bad for a sufficiency approach. But there might be a way out: distinguish causation involving omissions from "normal" causation, and insist that a sufficiency approach is well-suited to the latter, a dependency approach to the former. (Hall 2004 adopts this approach; Lewis 2004b disputes it.) That's not crazy: causation involving omissions typically fails, for example, to enter into the causal patterns of preemption and overdetermination that are so easy to construct when one is dealing with garden-variety causal processes, and that is some reason for thinking it needs a distinct kind of treatment. More importantly, evidence in favor of such a bifurcation emerges when we consider the alternative: adopt a dependency approach across the board. Now, the problem is not with accommodating the varieties of causation involving omission. It is rather that the dependency approach does so in a way which (i) directly conflicts with **Transitivity**; (ii) undercuts the strategies sketched above for handling late preemption.

The trip neatly illustrates the first of these problems. If Joey had not seen Suzy, he would not have run toward her. If he had not run toward her, he would not have tripped. If he had not tripped, the window would not have broken. A dependency approach will see a causal link at each step. If causation is transitive, it will follow that Joey's seeing Suzy is a cause of the window's breaking. Examples like this are embarrassingly easy to construct: the presence of some threat to E causes a response which prevents that threat, thus ensuring E's occurrence; but it ought not to follow that the presence of the threat was *among the causes* of E. (Observe that a fan of bifurcation has a natural response: **Transitivity** is correct, but only of garden-variety causation, *not* of causation involving omissions.)

Examples like this have led some authors to deny transitivity (see for example McDermott 1995). But doing so won't help with the second problem. Recall that there were two strategies for coping with late preemption that looked at all promising: the intrinsicness strategy and the de facto dependence strategy. But the kind of intrinsicness thesis needed to underwrite the first of these strategies is systematically violated in cases of causation involving omissions. This is perhaps clearest in cases of double-prevention: Event E depends on event C only because C blocks some threat to E. So this dependence is secured by virtue of facts *extrinsic to* the "process" (such as it is) connecting C to E: namely, the *existence of the threat*. For if there is no threat, there is no dependence of E on C, and certainly no causation of E by C. That fact directly conflicts with the claim that the causal structure of a process is intrinsic to it (even when the claim is spelled out with proper precision: see Hall 2003). The bifurcationist can say that **Intrinsicness** holds only of one *kind* of causation—the kind exhibited in, for example, cases of late preemption. She can therefore avail herself of the intrinsicness strategy for handling such cases. The unificationist cannot.

What of the other strategy: say that in cases of late preemption (as elsewhere), the cause is distinguished as that on which the effect de facto depends? Well, consider once again cases of threat-canceling. Event C occurs, initiating a threat to E, by way of intermediate event D. But C also triggers a response to this threat. C is not thereby a cause of E. But E *does* de facto depend on C: for, *holding fixed that D occurs*, if C had not occurred then no response to the threat from D would have occurred, and so E would not have occurred. We get the same unacceptable result we saw two paragraphs ago. (As a nice corollary, we can see that it was hasty of McDermott et al to blame *transitivity* for these results.)

So the state of play seems to be this: A sufficiency approach must distinguish two kinds of causation, and content itself with providing a theory of just one of them: namely, garden-variety causation that does not involve omissions. Its prospects for handling the most difficult examples—cases of late preemption—seem to me to be good; in doing so, it will necessarily place the **Intrinsicness** thesis at center stage. A dependency approach need not distinguish two kinds of causation, but cannot appeal to **Intrinsicness** in order to handle late preemption, and may be able to pursue the alternative de facto dependence strategy only at the cost of awkward verdicts about cases of threat-canceling.

§7 <u>The value of further work</u>

It's a tangled dialectical thicket, and it only gets worse once one starts to consider other varieties of preemption, symmetric overdetermination, different challenges to transitivity, issues involving the asymmetry of causation…. What's more, one needs to remember that, in the face of some horribly unintuitive consequence (e.g., that canceled threats turn out to be causes), it is open to a theorist to bite the bullet, and remind us all that the theoretical utility of the account may well pay for its lack of fit with intuition. It's safe to say that *a lot* of work remains to be done before we will be able to command a clear view of the best possible accounts, let alone of their distinctive theoretical virtues. I'll close by quickly considering one specific payoff that such work ought, in my view, to aim for.

I noted, in the introduction, that there has been in the last decade or so an upsurge of interest in causation *outside* of philosophy. 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 to my mind two serious problems as yet prevent this approach from attaining the kind of scientific rigor it ought to have. Both are foundational. First, 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. Second, the way causal models are typically used draws no distinction whatsoever between ordinary causal processes and causal connections involving omissions—so, for example, no distinction whatsoever is drawn between the way in which Suzy's throw causes the window to break, and the way in which Joey's trip causes the window to break. That seems to me a serious mistake. Earlier, in §5, I spent a lot of time shoring up nomological entailment accounts by criticizing their rivals. But I think a more inspiring reason for continuing to pursue them is that no other approach has, as far as I can see, a prayer of providing the kind of account of causation that can set this important area of research on a sound conceptual footing. Every once in a while the boundaries between philosophy and some or other scientific discipline blur, in a way that offers up to philosophers the chance to make a genuine contribution to their non-philosophical colleagues. Now is such a time.

§8 REFERENCES

Anscombe, G.E.M., 1971. Causality and Determination: An Inaugural Lecture. Cambridge: Cambridge University Press.

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.

Beebee, Helen 2004. “Causing and Nothingness”, in Collins, Hall, and Paul (eds) 2004.

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

Cartwright, Nancy 1999: The Dappled World, Oxford: Oxford University Press.

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.

Eells, Ellery 1991. Probabilistic Causality. Cambridge: Cambridge University Press.

Ehring, Douglas 1997. Causation and Persistence, New York: Oxford University Press.

Fair, David 1979. “Causation and the Flow of Energy”, Erkenntnis 14: 219-50.

Good, I. J. 1961. “A Causal Calculus I,” British Journal for the Philosophy of Science 11: 305-18.

Good, I. J. 1962. “A Causal Calculus II,” British Journal for the Philosophy of Science 12: 43-51.

Hall, Ned 2002a. “Non-locality on the Cheap? A New Problem for Counterfactual Analyses of Causation”, Noûs 36: 276-294.

Hall, Ned 2003. “The Intrinsic Character of Causation”, in *The Metaphysics Annual*, (Zimmerman ed.), Oxford Univ. Press.

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

Hall, Ned forthcoming. *Causation*, Oxford: Oxford Univ. Press.

Hall, Ned and L. A. Paul 2003. “Causation and Preemption”, in *Philosophy of Science Today* (Clark and Hawley eds.), Oxford Univ. Press.

Hall, Ned and L. A. Paul forthcoming. *Causation and the Counterexamples: A Traveler’s Guide*, Oxford: Oxford Univ. Press.

Hitchcock, C. 2001. “The Intransitivity of Causation Revealed in Equations and Graphs”, Journal of Philosophy 98: 273-299.

Hitchcock, C. ed. 2004. Contemporary debates in philosophy of science, Malden, MA: Blackwell.

Hume, David 1748. An Enquiry Concerning Human Understanding.

Kim, Jaegwon 1971. “Causes and Events: Mackie on Causation”, Journal of Philosophy 68: 426-41.

Kim, Jaegwon 1973b. “Causes and Counterfactuals”, Journal of Philosophy 70: 570-572.

Kvart, Igal 2004. “Causation: Probabilistic and Counterfactual Analyses”, in Collins, Hall, and Paul (eds), 2004.

Lewis, David 1970. “How to Define Theoretical Terms”, Journal of Philosophy 67: 427-446. Reprinted in Lewis 1983a: 78-95.

Lewis, David 1973a. “Causation”, Journal of Philosophy 70: 556-67. Reprinted with postscripts in Lewis 1986a: 159-172.

Lewis, David 1973b. Counterfactuals, Cambridge: Harvard University Press.

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 1986a. Philosophical Papers, Volume II, Oxford: Oxford University Press.

Lewis, David 1986b. Postscripts of “Causation”, in Lewis 1986a.

Lewis, David 1986c. “Events”, in Lewis 1986a: 241-269.

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.

McDermott, Michael, 1995. “Redundant Causation”, British Journal for the Philosophy of Science 46, 523-544.

McGrath, Sarah, 2002. “Causation by Omission”, Chapter 2 of Sarah McGrath Causation in Metaphysics and Moral Theory, Ph.D. dissertation, MIT.

Mellor, D.H. 1995. The Facts of Causation, London: Routledge.

Menzies, Peter 1996. “Probabilistic Causation and the Pre-Emption Problem”, Mind 105: 85-117.

Pearl, Judea 2000. Causality: Models, Reasoning and Inference. Cambridge: Cambridge University Press.

Salmon, Wesley 1980. “Probabilistic Causality,” Pacific Philosophical Quarterly 61: 50 - 74.

Salmon, Wesley 1994. “Causality Without Counterfactuals”, Philosophy of Science 61: 297-312.

Sankey, Howard (ed) 1999. Causation and Laws of Nature, Dordrecht: Kluwer.

Schaffer, Jonathan 2004. “Causes need not be Physically Connected to their Effects: The Case for Negative Causation”, in Hitchcock (ed.) 2004. Shoemaker, S. 1980. “Causality and Properties”, in van Inwagen 1980: 109-35.

Thomson, Judith 2003. “Causation: Omissions”, Philosophy and Phenomenological Research 66, 81-103.

Tooley, Michael 1987. Causation: A Realist Approach, New York: Oxford University Press.

Tooley, Michael 1990. “Causation: Reductionism versus Realism”, Philosophy and Phenomenological Research 50, Supplement: 215-36.

van Inwagen, Peter (ed) 1980. Time and Cause, Dordrecht: Reidel.

Yablo, Stephen 2002. “De Facto Dependence”, Journal of Philosophy 99: 130-148.