

OUTLINE OF A DISPOSITIONAL ACCOUNT OF CAUSATION

Abstract

I outline and consider the prospects for an account of causation suitable for a metaphysics of dispositions. Such an account suggests that causes are conditionally sufficient (rather than necessary) for their effects. I consider what semantics for subjunctive and counterfactual conditionals this requires, and examine the consequences for making a distinction between causes and mere conditions and between collective or complete causes and mere parts of causes.

1 Introduction—dispositions and causes

What is the ontological ground of causation? That is to say, what, in general terms, are the features of the world on which the existence of causal relations depends? For Lewis (1973) it is the laws of nature, via the role they play in fixing the similarity relations amongst worlds and so the truth of the counterfactuals. For others too, such as Armstrong (1983), Dretske (1977), and Tooley (1977), it is also the laws of nature. But for yet others laws are not fundamental relative to causation in this way, since the ground of laws, if they exist at all, is causal power (Swoyer 1982, Ellis 2001, Mumford 2004, Bird 2005). According to such a view nomic and causal relations supervene on the existence of properties that are dispositional in nature.

Adherents of this view of the ground of causation and laws need not say anything special about causation. They could, if they wished, maintain that essentially dispositional properties account for the laws of nature, a very non-Lewisian view, but then agree with Lewis as regards the analysis of counterfactuals in terms of laws and the proximity of possible worlds and as regards the analysis of causation in terms of counterfactuals.

However, once one focuses on dispositions, it is natural to think that there ought to be some more direct connection between dispositions and causation. A flammable liquid is ignited, and catches light. The igniting of the liquid is the cause, and it is tempting to think that the flammability of the liquid plays some role here: it is because the liquid is flammable, i.e. disposed to catch light when ignited, that the igniting of the liquid can cause it to catch light. This suggests a simple dispositional analysis of causation:

(SD) c causes e iff e is the manifestation of a disposition of which c is the stimulus.¹

¹Jonathan Jacobs makes such a proposal.

In what follows I shall investigate the prospects for (SD) by considering the connection between dispositions and subjunctive conditionals. This has three benefits. First, 'disposition' is to some degree a term of art, and our reasoning and intuitions are better developed with conditionals than with dispositions. Secondly, as I shall briefly mention shortly, there are ontological issues associated with the term 'disposition' that are important but which it is best to set aside for consideration later. Thirdly, by discussing the dispositionalist proposal in terms of conditionals, we will be able to see more clearly the contrast between this view and Lewis's counterfactual conditional account.

Let us turn then to the link between dispositions and conditionals. We'll start with the simple conditional analysis of dispositions:

(CA) x is disposed to yield manifestation m in response to stimulus s iff were x to receive stimulus s it would yield manifestation m .

I now make two important caveats, concerning simplifications contained in (SD), (CA), and our use of them:

(I) This analysis is subject to certain counterexamples (Johnston 1992, Martin 1994, Bird 1998). Nonetheless, for now it will be instructive to employ the simple conditional analysis as it stands. In due course we will return to the counterexamples, to see what impact they have on the analysis of causation provided.

(II) We will need to consider how strongly we should take the ontological implications of 'a disposition' in (SD). For the time being we will take a liberal view: something's being disposed a certain way, as given by the conditional in (CA), suffices (and is necessary) for the existence of a disposition. Nonetheless, our use of (CA) should *not* be taken to imply that conditionals are the ground of dispositions. The relationship is, rather, the other way around. But if dispositions ground conditionals, we will need ultimately to take them seriously in ontological terms. Dispositions may not be as ubiquitous as (CA) suggests.

These simplifications should be regarded as akin to the use of an idealized model in science to which we will later take into account known deviations from the ideal. To the extent that taking such deviations into account improves the explanatory fit of the model, the model receives additional confirmation.

Putting (SD) and (CA) together we have the following account of causation:

(SD-CA) c causes e iff for some x , were x to receive stimulus c , then x would yield e , c occurs, and e occurs.

The existential quantification 'for some x ' plays only a dummy role. If the counterfactual ' c occurs $\square \rightarrow e$ occurs' is true, then the world as a whole is something such that were it to receive stimulus c it would yield manifestation e . (This is something to which we will return when considering caveat (II).) Consequently, we can simplify (SD-CA) to:

(SD-CA)' c causes e iff c occurs $\Box \rightarrow e$ occurs $\wedge c$ occurs $\wedge e$ occurs.

(SD-CA)' needs careful explanation and defence, which it will receive in the following sections. For present, ' $\Box \rightarrow$ ' symbolizes, as is usual, a certain species of conditional whose instances are often counterfactual conditionals. But not all are; counterfactual conditionals in English imply, pragmatically at least, the falsity of their antecedents (as their name suggests). But we do not take ' $p \Box \rightarrow q$ ' to imply the falsity of p . The conditionals in question should preferably be called *subjunctive* conditionals, since that is the grammatical mood typically employed in the counterfactual case ('had you gone out now, you would have got wet') and the non-counterfactual case ('were you to go out now, you would get wet'). In what follows the relation symbolized by ' $\Box \rightarrow$ ' will often be such a non-counterfactual subjunctive conditional.

(SD-CA)' makes causes conditionally sufficient for their effect. ' $\Box \rightarrow$ ' involves an implicit reference to background conditions that are relevantly similar to actual conditions. ' c occurs $\Box \rightarrow e$ occurs', says that given such conditions, if c occurs then e occurs; that is, in such conditions, the occurrence of c suffices for the occurrence of e . This conditional sufficiency I call 'subjunctive sufficiency' and contrasts, of course, with Lewis's claim that causation amounts to counterfactual necessity. According to Lewis, causation between c and e requires ' c does not occur $\Box \rightarrow e$ does not occur' (or its ancestral), i.e. causes are conditionally necessary for their effects.

In the remainder of this paper I shall develop (SD-CA)' to see how far it can be taken as an account of causation. I shall then consider objections to (SD-CA)'. We will see that to a large degree these can be accommodated by revisiting the caveats (I) and (II)—that is the objections can be explained away as arising from the counterexamples to (CA) and to failure to take the ontology of dispositions sufficiently seriously. To that extent (SD) remains intact. I shall conclude by considering the extent to which (SD) provides an illuminating account of causation.

2 The denial of centering

One immediate problem with (SD-CA)' is that it seems to have the consequence that any actual event is the cause of any other actual event. According to Lewis's account of counterfactuals the following holds:

(C) $A \wedge B \rightarrow A \Box \rightarrow B$.

(C) and (SD-CA)' entail:

$(c \text{ occurs} \wedge e \text{ occurs}) \rightarrow c \text{ causes } e$.

And so any two actual event are causally related.

(C) is the (strong) centering condition. In terms of Lewis's possible worlds semantics for counterfactuals, (C) is the claim that no world is as similar to the actual world as the actual world itself. A defender of (SD-CA)' must reject (C).²

²Rejecting centering does not entail rejecting weak centering, which is the denial that there is a possible world *more similar* to the actual world than the actual world itself. That said, reasons for rejecting centering might well lead us to reject weak centering also. Cf. Gundersen 2004: 12–13.

There are indeed good reasons to reject centering. In the current context, one is the fact that (C) and (CA) together entail that event every two actual events are dispositionally related. One might regard that as further proof that (CA) is defective. But I think that would be a mistake. Consider two scientists discussing whether some experimental sheet of glass is fragile or not. A says, “The glass is fragile; if I were to strike it, it would shatter”. “No,” says, B, “This is really good stuff, I promise you. It isn’t fragile; even if you were to strike it fairly hard, it wouldn’t shatter”. “Well then,” says A, “Let’s see.” And so saying A takes up a hammer and strikes the sheet of glass with moderate force. Now consider two scenarios. In scenario I the hammer strikes the glass and the glass shatters. In scenario II, just as A is striking the glass, an enormous explosion in another laboratory violently rocks the building, causing the glass to shatter just after the hammer makes contact with the glass.

We would probably regard scenario I as vindicating A’s claims that the glass was fragile, and that it would shatter when struck. But we would not regard scenario II as vindicating A. Nonetheless, A, even in scenario II, can claim that according to (C) he was correct at least in saying that the glass would shatter if struck. But the naturalness of their conversation shows that such a response is mistaken. The explosion ruined their experiment. Nothing said by either A or B was confirmed or refuted in scenario II. In which case (C) is erroneous

Note, of course, this means that the outcome in scenario I is not an irrefutable confirmation of A’s claims. After all, there *may* have been some hidden process in scenario I that operated like the explosion in scenario II to cause the glass to shatter quite independently of the striking. Since the scientists did not notice anything of the sort, they can regard the experiment as pretty strong confirmation of A’s claims. But in analogous experiments we may not be so confident. Replication may reduce the probability that that the outcome was not the result of the intervention but instead of some hidden independent process. A control, if available, will often serve that purpose better. Had the scientists had a second piece of glass that was not struck, the fact that it shattered in scenario II would have shown them that the shattering of the struck glass tells them nothing about the truth of the subjunctive ‘were it struck, it would break’. Correspondingly, such a control in scenario I would raise their confidence that no hidden mimicking process is at work.³

This concern to eliminate mimickers is another manifestation of our disinclination to accept (C). If we accepted (C) we would not be concerned to eliminate mimickers, since a mimicking process would make the relevant counterfactuals and subjunctives true. Before continuing I should address the worry that it is not (C) that is at the root of our concern, but (CA), which we already know to be faulty—and thanks to mimickers and other things of that sort. Such a worry would have us drive a wedge between the two parts of A’s claim, ‘The glass is fragile; if I were to strike it, it would shatter.’ In scenario II the subjunctive ‘if I were to strike it, it would shatter’ is shown to be true; but because (CA) is false, it does not follow that the glass is fragile. As mentioned, the naturalness of the conversation between A and B and the unnaturalness of taking scenario II to vindicate the subjunctive claim, suggest that

³A mimicking process is one which brings about the same manifestation, e.g. shattering, but independently of the supposed stimulus, e.g. the striking. Cf. Johnston 1992.

(CA) has more going for it intuitively than (C). But at this point I want to address the suspicion that the very same reasons we glean from a more detailed consideration of (CA) and which lead to its rejection, are at play here.

Consider scenario III, in which scientist A has attached a small explosive device to the same sheet of glass. That device has a very sensitive detonator. A strikes the glass, the striking disturbs the detonator, and the resulting explosion shatters the glass. In this case B is forced to concede, once she realises what has happened, that A's subjunctive 'if I were to strike it, it would shatter' has been verified. But B may reasonably reject A's claim that the glass is fragile. Thus scenario III does show that (CA) is false,⁴ and at the very least needs modification. But notice the different role of mimicking in the two cases. Consider

- (a) c occurs \wedge e occurs;
- (b) c occurs $\Box \rightarrow e$ occurs;
- (c) there is a disposition to manifest e in response to c .

In scenario III we have an entirely kosher subjunctive conditional, but that conditional mimics the disposition. Whereas in scenario II, the conjunction of the striking and the (independently) shattering glass mimics the conditional. In the former case we thus accept the conditional but deny the disposition; that is the correct description of scenario III denies the inference of (c) from (b), i.e. denies the truth of (CA). In describing scenario II, we accept the conjunction but deny the conditional, i.e. we deny the inference of (b) from (a), i.e. we deny (C). Thus the falsity of (CA) plays no part in our rejection of (C).

Other philosophers have pointed to the more general counterintuitive consequences of (C). Alan Hájek (2007: 46-8), after remarking that most people would be puzzled by rather than assent to counterfactuals joining unrelated actual facts, such 'If Canberra were the capital of Australia then the moon would have large craters', goes on to point out that matter are worse when the events are related but in such a way that the antecedent reduces the chances of the consequent. Furthermore, since ' $A \wedge B$ ' is symmetrical for A and B, $A \wedge B$ entails not only $A \Box \rightarrow B$, but also $B \Box \rightarrow A$. So both of the following are true of the 2000 U.S. presidential election: 'if Gore had won the popular vote, then Bush would have won the election' and 'if Bush had won the election, then Gore would have won the popular vote'.⁵

The rejection of centering means that when A and B are actually the case, more worlds are relevant to the truth of $A \Box \rightarrow B$ than just the actual world. Which additional worlds are these? Nozick, who needs a subjunctive conditional that denies (C), takes the relevant worlds to be the set of *nearby* possible worlds, that is the set of worlds closer (more similar) to the actual world than some threshold distance. Gundersen

⁴At least if we assume: x is fragile $\equiv x$ is disposed to shatter when struck.

⁵Hájek also discusses cases that turn on the incompatibility of 'If x had occurred, y might not have happened'/'If x had occurred, y would have happened'. E.g. I am about to toss a fair coin. I say truly, 'if I were to toss this coin, it might land tails'. I do toss the coin and it lands heads. Thus according to centering 'if I were to toss this coin, it would land heads' would have been a correct thing to say. But the truth of the latter is incompatible with what I did say. For an extended discussion of centering and an alternative semantics for counterfactuals, based around *normality* rather than *similarity* and which rejects centering and weak centering, see Gundersen 2004.

takes the worlds to be those that are *normal*. Taking the latter route means the rejection of weak centering—since abnormal things do occur, the actual world may not be among the normal worlds in certain respects. Hence the set of worlds relevant to the truth of $A \Box \rightarrow B$ would not only include worlds other than the actual world, but need not include the actual world at all. One might of course regard the relevant set as intersection of the sets of nearby and normal worlds (or one's conception of normality might imply nearness, even if not implied by it). For current purposes I shall take the nearby worlds to be those relevant to the truth of subjunctive conditionals, although I shall also consider the benefits of adding the restriction to normal worlds also.

3 Causes and conditions

Lewis, along with many other philosophers who take counterfactual necessity as the mark of causation, draws no ontological distinction between a cause of an effect and a condition for that effect's occurrence. Yet, this is a distinction that non-philosophers find it natural and easy to make.⁶ In an ordinary case of a struck match lighting, every non-philosopher takes *the* cause of the lighting to be the striking and will consign the presence of oxygen to a lower status. Flicking the switch caused the bulb to illuminate, not the continued functioning of the power station supplying the electricity. The distinction between a cause and a causally necessary, background condition is even more obvious the further back we trace such conditions. A person's birth is a necessary condition of their death at aged 82 of a heart attack, but their attempting to run a marathon was its cause.⁷

Those who think that causes are counterfactually necessary conditions must regard all of these events as causes. For Lewis the distinction is merely pragmatic—all these events are indeed causes of the events for which they are necessary conditions. But our interests may lead us to focus our attention on one of the many causes and to pick it out as *the* cause.

We sometimes single out one among all the causes of some event and call it “the” cause. Or we speak of the decisive or real or principal cause ... I have nothing to say about these principles of invidious discrimination.
(Lewis 1986: 162)

Likewise Hall (2004) sermonizes (his term),

Suppose that my favourite analysis counts the Big Bang as among the causes of today's snowfall... How easy it is to refute me, by observing

⁶The advantage of this account in making the cause–condition distinction was made clear to me by reading Broadbent (2007a, 2008). Broadbent himself proposes a ‘reverse counterfactual’ account of causation, where the key counterfactual is: $\neg e \Box \rightarrow \neg c$ (see also Broadbent 2007b). This is a sufficiency account of causation and so, like mine, generates the cause–condition. Below I shall consider the relative merits of Broadbent's view and mine.

⁷Those of Humean inclinations may not like this example, since birth is a logically necessary condition of death and so this may be thought to introduce an illegitimate necessary connection between distinct events. In which case we may substitute for a person's birth, their being fed as a child or their being administered penicillin during an early illness.

that if asked what *caused* the snowfall (better still: what was *the* cause of it), we would never cite the Big Bang! Of course, the right response to this “refutation” is obvious: It conflates the transitive, egalitarian sense of “cause” with a much more restrictive sense (no doubt greatly infected with pragmatics) that places heavy weight upon salience.

Hall’s response is disingenuous. What reason do we have for supposing that there is an egalitarian sense of cause that encompasses the Big Bang as a cause of every particular event, and my grandmother’s birth as a cause of all my actions, and so forth—except for the fact that counterfactual accounts such as those of Lewis and Hall deliver that result? Elsewhere our intuitive reactions to certain cases are regarded as data that ought if possible be accommodated by a satisfactory theory. But here those reactions are dismissed as infected by pragmatics.

The resort to pragmatics is an insufficient response. For a start, as Broadbent rightly complains, those who make it never offer a substantive account of the pragmatic principles at work. And when we turn to our best general account of the pragmatics of discourse, Grice’s account of conversational implicature, we find that Grice’s principles do not deliver the result that is required, as Menzies shows. Moreover, salience and pragmatic concern may well focus one’s interest on a necessary condition without thereby elevating to the status of cause. A speeding motorist causes an accident. The cause of the accident is clearly their excessive speed. Even so, one might take an interest in other factors. The town council may conclude that any of a variety of speed reduction devices (speed humps, road narrowing, speed cameras, etc.) would have prevented the accident. But that does not mean that the council concedes that the lack of humps, the width of the road, or the absence of speed cameras are each causes of the accident. In which case the application of the honorific ‘cause’ is not correlated with the focus of our interests.

Lewis in effect has an error theory of our normal causal talk, since ‘*the* cause’ implies only one cause, whereas according to Lewis there is never only one cause. While such an error theory could be right, a view of causation that respects our intuitive distinction between cause and condition will have a significant advantage. The dispositional account respects that distinction. Consider the lighting match. Clearly the unstruck match has a disposition to light in response to the stimulus of being struck. But it does not have the disposition to light in response to the stimulus of being in the presence of oxygen—not under these circumstances. The elderly gentleman was disposed to die of a heart attack in response to excessive exertion, but not disposed to die of a heart attack in response to being born (or being fed as a child etc.).

The point is most readily appreciated by looking at (SD-CA)’. Given our dropping of centering, this requires that the material conditional, $c \rightarrow e$, holds in nearby worlds as well as in the actual world. Consider $c \equiv$ the match is struck, and $e \equiv$ the match lights. In some nearby worlds the match may not be struck, in which case $c \rightarrow e$ is true. In others it is struck, but since in nearby worlds oxygen is present the match lights, so again $c \rightarrow e$ is true. The striking of the match causes its lighting. But now consider $c' \equiv$ oxygen is present. In some nearby worlds the match is not struck, but oxygen is still present. Hence in those worlds $c' \rightarrow e$ is false, and so $c' \Box \rightarrow e$ is also false.

The presence of oxygen is not a cause of the lighting. Likewise, *mutatis mutandis*, for the other cases.

4 Collective causes

The dispositional account has the following feature. If two uncommon events coincide to cause an effect, so that both are required—neither is sufficient in the absence of the other—then neither counts as a cause of the effect. For example, a fire occurs because a fuel line leaked and a build-up of static electricity (perhaps due to unusual weather conditions) caused a spark. The set-up was not disposed to bring about a fire either in response to the fuel leak or in response to the spark, but only in response to both together. So neither individually causes the fire. In counterfactual terms, [there is a leak $\square \rightarrow$ there is a fire] is not true, since in a nearby world there is no spark and so no fire, and likewise [there is a spark $\square \rightarrow$ there is a fire] is not true since in a nearby world there is no fuel leak and so, again, no fire.

This might be thought to be a disadvantage of the dispositional account. Surely the fuel leak and the spark are both causes? I do not think so. I suggest that philosophers may be inclined to think so because of so lengthy an exposure to the Hume–Lewis negative counterfactual view which make any necessary condition a cause. But as we have already seen, everyday causal talk is not so profligate with ascriptions of cause, making a distinction, as it does, between causes and mere conditions. In this case too, everyday causal talk is disposed *not* to take the individual events as causes but rather takes them together as a collective cause. In answer to the question, ‘what caused the fire?’ it would generally be regarded as incorrect, rather than at worst misleadingly incomplete, to say just ‘the fuel leak’ or just ‘the spark’. No, what caused the fire was the fact that both these events occurred together. This answer is most clearly correct when the two (or more) components of the cause are identical. What caused Oedipus to be blind? Not that he removed his left eye, nor that he removed his right eye, but that he removed both eyes. A twin-prop aeroplane crashes as a result of the failure of both of its tow engines: the cause of the crash is just that, the failure of both engines, not the failure of either one of them. I fill my tank with 50l. of fuel so as not to have to stop on my car journey. The first 25l. that I put into the tank does not enable (‘enable’ = ‘causes to be possible’) me to get to my destination, nor does the second 25l., only that I filled with 50l. Fraser MacBride, in a rather different context, also appeals to the idea of collective causation, in his example of an unfortunate man stung to death by a swarm of bees: no individual bee sting is the cause of his death; rather the many stings are collectively the cause of death.⁸

And these are the verdicts that the dispositional account yields. As mentioned, the state of affairs before the fire was disposed to bring about a fire as a result of the co-occurrence of a fuel leak and spark, but not in response to either of those events individually. In counterfactual terms, [there is a fuel leak and there is a spark $\square \rightarrow$ there is a fire] is true. Likewise [the left engine fails $\square \rightarrow$ the aeroplane crashes] is false,

⁸MacBride [“The Particular–Universal Distinction: A Dogma of Metaphysics?”, *Mind* 114 (2005) 565–624] is concerned to show that causation is a multigrade relation.

likewise [the right engine fails $\square \rightarrow$ the aeroplane crashes], but [the left engine fails and the right engine fails $\square \rightarrow$ the aeroplane crashes] is true. Similarly, the first 25l. and the second 25l. may both be individually necessary for my travelling without needing to stop, but only the full 50l. is sufficient.

5 Counterfactual dependence?

Ned Hall claims that there are two concepts of causation, which he calls *dependence* and *production*. Dependence is negative counterfactual dependence—counterfactual necessity. Hall offers a tentative analysis of production, a concept that is closer to some kind of sufficiency. In Hall's view, to account for all our causal ascriptions we need both of these distinct concepts. I suggest that we can do without dependence and that production should be understood in terms of counterfactual sufficiency.

Hall considers someone who might react to his discussion by denying that counterfactual dependence is causation. The defender of counterfactual dependence says,

Nonsense; counterfactual dependence *is too* causation. Here we have two wholly distinct events; moreover, if the first had not happened, then the second would not have happened. So we can say—notice how smoothly the words glide off the tongue!—that it is in part *because* the first happened that the second happened, that the first event is partly *responsible for* the second event, that the occurrence of the first event helps to *explain why* the second event happened, etc.

Note that Hall, says 'in part' and 'partly' and 'helps'. So it seems that even Hall is uncomfortable with saying *a* causes *b* when *a* is just one among several necessary conditions (and hence not sufficient). Indeed, everything Hall says above, except the first sentence, is consistent with the denial that the first event causes the second event. The best explanation of the natural use of the qualifiers 'partly' and the like is that the events in question are *parts* of causes—how else should we interpret '*b* partly because *a*' except as 'for some *c*, *b* because *c* and *a* is part of *c*'.

Since I reject dependence as a component of an analysis of causation, I am under some obligation to offer an explanation of why it should seem to many that it is a (or *the*) central feature of causation. The fact that dependence will pick out parts of causes is itself one reason why we might mistakenly fix on dependence as an account of causation. If dependence or an account based on it (such as Lewis's) were, by the lights of its supporters, to be a satisfactory account, then it would differ in its evaluations from the dispositional view only insofar as it also counts as causes (a) parts of causes) and (b) conditions. We have discussed conditions above. Why theorists have been willing to swallow this counterintuitive aspect of the dependence view is in part testament to the power of theory to inform our or override our intuitive judgments. This may be helped by misleading examples that seem to suggest that what we regard as a cause and what we regard as a condition is interest relative. So while we may be think that the presence of oxygen is a mere condition for the fire, not a cause, we may be persuaded to think otherwise by relating a story in

which a seemingly similar set of events occurs in which we undoubtedly do identify the oxygen as a cause. If oxygen were normally absent from some process, but was accidentally introduced and as a result a fire occurred, then the oxygen would be a cause. But that does not show that it was a cause in the case where it *is* normally present. The two cases differ with respect to their dispositions and to the truth of the positive subjective conditionals, and so the difference in our intuitive judgment is not explained by a difference in our interests.

Another reason why dependence seems important is that it is related to responsibility. Human nature being what it is, we are very interested in who can be responsible for those occurrences that we would rather had not happened—rather more interested than in finding out who brought about events that we are happy with. As such responsibility is a matter of necessary conditions—one is responsible for what one could have prevented but failed to. And it may be natural to think that what we are doing by identifying those persons and actions that are responsible for a certain outcome is identifying the cause of the unfortunate events in question. In which case we will equally naturally associate causation and necessary conditions, and given that association we can generalize beyond cases of human action and responsibility to relations between events in general

Persons and actions that are responsible for certain outcomes will often be causes of them. This is because intentional behaviour is typically both a necessary and a sufficient condition of the intended outcome. Lucy wants her garden to be watered. On a summer's day she will want to bring about that outcome, i.e. cause it, and so engage in an action that is subjunctively sufficient for the garden to be watered. On the other hand, if the desired outcome is going to happen anyway, and we know that, then we have no reason to act, and human nature being what it is we typically desist from acting. If Lucy can see that it is about to rain she won't go to the effort of getting the hose out. So our actions are typically not only sufficient conditions of their outcomes but also necessary conditions—we often do not act unless it is necessary to do so. This means that actions that are causes of their effects are also often necessary conditions of those effects and so it is not surprising that responsibility, causation, and necessary conditions are all lumped together. But causation and responsibility can come apart, as cases of negligence show. A landlord's negligently failing to check the safety of his building may be a necessary condition, but not a sufficient one, of the fire that broke out. That fact may entitle us to hold him responsible, but we do not have to regard his negligence as a cause of the fire.

6 Problems and possible solutions

Having given a sketch of a dispositional account of causation above, I shall now turn to potential objections. The discussion so far has concentrated on subjunctive sufficiency, which given (CA) is a consequence of (SD). But, as remarked, (CA) is false. It is approximately true, but in some cases dispositions and subjunctive conditionals come apart. Given that they do come apart, we may ask, which of (SD) and (SD-CA)' is true. As we shall see, certain objections to (SD-CA)' disappear once we focus on

(SD). The account being proposed here is thus a dispositional rather than subjunctive account.

(a) If **a** were crimson, **a** would be red, and **a** is in fact crimson. But we do not regard **a**'s being crimson as a cause of its being red. Counterfactual dependency views of causation have a similar problem: if **a** were not red, it would not be crimson, which, on their view suggests that being red is a cause of being crimson (which is even less plausible than being crimson causing being red). However, the dispositional view can avoid this problem by remarking on the difference between dispositions and counterfactuals. We do not think that everything is such that it has the *disposition* to be red if it is crimson. So such a case already constitutes a counterexample to the counterfactual analysis of dispositions, (CA). Rejecting (CA) and so (SD-CA)' but retaining (SD) allows one to say that although it is true that if **a** were crimson, **a** would be red and that **a** is in fact crimson and red, there is no causation here because there is no disposition—this is one of the cases where dispositions and subjunctive conditionals come apart.

(b) Let it be the case that in the circumstances the only event that could cause *e* is *c* and that *c* did cause *e*. One might be inclined to say that it is true to say that were it that case that *e* occurs, then it is also the case that *c* occurs. Thus our subjunctive conditional account has the consequence that not only does *c* cause *e* but also that *e* causes *c*. Note that the counterfactual dependence view also suffers from this problem. Lewis's answer is to outlaw backtracking counterfactuals. If that answer is a good one, with independent motivation, then I can use it too. But if it is not, then the dispositional account can help without appealing to the denial of backtracking, since in normal cases there will be a disposition to yield *e* in response to *c* without there being a disposition to yield *c* in response to *e*. (I am not saying that we must always rule out backwards directed dispositions, just that generally they do not occur, which suffices to make the distinction in the direction of causation that such cases require.)

(c) If *c* is subjunctively sufficient for *e* then *c+d* is subjunctively sufficient for *e*, where *d* is causally independent of *e*. So spurious events may seem to be parts of causes where they are not. This is the rough analogue of the problem for Hempel that any condition can be added to laws and conditions that entail the explanandum without removing that entailment. So we get spurious explanatory and spurious causal factors. Again the appeal to dispositions removes the problem, since a fragile vase has the disposition to break in response to being stressed, but it does not have the disposition to break in response to the complex stimulus [being stressed and the U.S. President is a Republican].

Objections (a)-(c) reinforce (SD) by showing that objections against (SD-CA)' occur precisely where (CA) is false and so where (SD) and (SD-CA)' come apart. These are exceptions that prove the rule, and confirm (SD) in the way that observations in science support an underlying theory when we find that although they are inconsistent with a simple model built on that theory they are consistent with more sophisticated models constructed on the same basic theory (for example, in Newtonian mechanics, observations of the motion of the moon refuted his own model of the moon's motion, but were found eventually to be in conformity with a mathematically more sophisticated model. Those observations thus confirmed the underlying

Newtonian gravitational theory.) The next objection, however, pulls in the opposite direction.

(d) The occurrence of c may cause the occurrence of e , even though c and e are somewhat remote, and so do not seem to be related as stimulus and manifestation of a single disposition. For example, I knock over a cup of tea; the tea runs between the floorboards, and into a recessed light in the ceiling below; the liquid causes the circuit to short and the fuse in the fusebox to trip. Many would be happy to say that the spilling of the tea caused the fuse to trip. But is there any disposition that is the disposition for the fuse to trip in response to the spilling of the tea? If so, what object has that disposition? In Section 1, I adopted a liberal approach to the existence of dispositions: I said that if the counterfactual ' c occurs $\square \rightarrow e$ occurs' is true, then the world as a whole is something such that were it to receive stimulus c it would yield manifestation e . So that approach would license, the causal claim via (SD-CA)'. But we have just seen that the appropriate responses to objections (a)-(c) require us to take a *less* liberal approach to the existence of dispositions, and to deny that there is a disposition for every true subjunctive conditional. Taking a more ontologically conservative, robust approach to dispositions allowed us to deny, for example, that there is a disposition to break in response to the complex stimulus [being stressed and the U.S. President is a Republican], or disposition to be red in response to being crimson. But this same ontological robustness seems likely to rule out a disposition in the case of the spilt tea also.

A plausible response to the spilt tea case, while observing a requirement for ontological robustness, is to maintain that a complex disposition is admissible for the purposes of (SD) only if it is constructed out of a chain of dispositions that are themselves ontologically robust. While this still leaves the notion of an ontologically robust disposition undefined, it provides an answer to the current problem. If a fuller articulation of the notion of an ontologically robust disposition is required, one might start with the idea of a primitive ontologically robust disposition, i.e. a disposition that is a fundamental natural property, and then regard an ontologically robust disposition as one that is built up from the primitive cases in a recursive manner. It is not my intention to defend such a view here, but merely to point out that it solves the spilt tea case without falling back on a notion of disposition weakened sufficiently to reintroduce problems (a)-(c).

The thought that a causal relation is often really a causal processes made up of a chain of several simpler causal relations is a natural one. It is also one that is present in Lewis's refinement of the counterfactual dependence view. However, I do not think that the naturalness of this idea supports both views equally. In my case, the appeal to causal chains comes about precisely because of the complexity of many causal relations, and so looking at the part of the complex relation is a natural response. But it is *not* complexity that drives Lewis to look at causal chains, but rather a certain kind of counterexample—pre-emption cases. So, on the one hand quite complex processes can satisfy a simple counterfactual analysis without seeming to call for being considered as causal chains, while very much simpler cases are counterexamples to the same simple analysis. So when we consider the famous case of Suzy and Billy throwing stones, it is not the complexity of Suzy's throw that suggests that we should look to causal chains, but rather the fact that Billy threw also. After

all, when Suzy throws alone—exactly the same causal process—there is no problem with the counterfactual analysis, and there is no urge for us to think in terms of causal chains. So although both the counterfactual dependence view and the dispositional view of causation invoke causal chains, they do so for different reasons—for the counterfactual view to deal with pre-emption cases, and for the dispositional view to deal with complex causal relations to which no simple disposition corresponds. The latter reason matches our intuition that causation is often complex, whereas the former reason does not.

(e) The immediately preceding discussion notwithstanding, the pre-emption problem does have an analogue for the subjunctive sufficiency view. Our final objection points out that two or more events may be subjunctively sufficient for some effect, but only one of them is the cause. For the counterfactual dependency view, the problem is that neither putative cause comes out as a cause, even though one clearly is. For the subjunctive sufficiency account, the problem is that both do, even though one clearly isn't. This is the analogue of Achinstein's poisoning problem for Hempel's D-N model of explanation. Thus Jones consumed a pound of arsenic, which is sufficient to kill him within 24 hours. But he in fact was killed by being hit by a bus. And, clearly, being hit by a bus was sufficient to kill him also.⁹ One might start by pulling the same trick as before, by moving to dispositions. His death was the manifestation of the disposition to die in response to a large impact, not a manifestation of the disposition to die in response to being poisoned. But as it stands that response needs further elaboration. He certainly has the disposition to die in response to being poisoned. Furthermore he was actually poisoned and he did indeed die. So the susceptibility to poison disposition did receive its characteristic stimulus and the possessor of that disposition did undergo an event that is the characteristic manifestation of the disposition. What makes it the case that Jones's death was not the manifestation of the susceptibility-to-poison disposition?

The pre-emption problem motivates the appeal to causal chains for the counterfactual dependency view, in a way I have suggested is unsatisfying. However, the dispositional view has causal chains for better reasons, and the causal chain idea may be employed here also. Indeed Lewis's solution translates directly to one for me. If we can see that some intermediate event, d , depends counterfactually on the predecessor event c_1 in the pre-empting chain but not on any event c_2 in the pre-empted chain, then it follows that c_2 is not subjunctively sufficient for d . It is clear that in such cases there is no chain of activated dispositions along the pre-empted chain that leads all the way to the effect.

Late pre-emption cases, however, make difficulties for both accounts that cannot be easily resolved by resort to the causal chain idea. Let us imagine that the causal chains in question can be reduced to a finite sequence of events: the pre-empting chain, whose final event is the effect e and whose penultimate event is c_1 , and the preempted chain whose final event is c_2 . One, ultimately unsatisfactory, answer is to endorse a conception of event identity that is sufficiently fragile, so that

⁹As far as I am aware, the fact has not hitherto been pointed out, that this case offers not only a counterexample to the D-N model of explanation (by making the poison explain the death) but also to the counterfactual dependency view of causation (by denying that

in the absence of c_1 , e would not have occurred, even if a very similar event would have occurred. The Lewisian answer is to appeal to a notion of causal influence—counterfactual dependence is causal when alteration in the antecedent event leads to alteration in the consequent event. But it is at best unclear whether this answer can deal satisfactorily with all late preemption cases, including cases of trumping.

This is a problem if we are attempting to characterise causation in terms of some much more general relation, counterfactual dependence. It is equally a problem if we are trying to characterise causation in terms of subjunctive sufficiency. But that is not what this project is. Subjunctive sufficiency is a rough and ready proxy for the relation between the stimulus and manifestation of a disposition. We have seen how rough it is. So focus has shifted onto that relation itself. The conclusion of this part of the discussion is straightforward. Remember that dispositions relate event (or fact/condition) *types*. Let s and m be token events of the S and M types. Then $\langle s$ occurs, a is disposed to manifest M in response to S, and m occurs \rangle does not entail that m is the manifestation of the disposition—it might be the manifestation of some other disposition with the same manifestation type. In the case of some macro-dispositions or macro-processes we may be able to distinguish between the real cause (the true stimulus, whose simulation of an appropriate disposition gave rise to the manifestation) and a pseudo cause (an event s in the presence of some disposition to manifest M in response to S), by investigating the causal chain—that is what we can do in the Achinstein poisoning case. But for manifestations of basic dispositions that will not be possible. We must simply say that the stimulus-disposition-manifestation is primitive.

One might worry that this answer is unsatisfying. We were looking for a revealing analysis of causation and we are ultimately given something that doesn't look so very far from causation and which if anything is rather more obscure. To reiterate, whether one ought to find that unsatisfying rather depends on what one was looking for in the first place. If one was looking for some unpacking of the concept of causation in terms of more familiar concepts, or in terms, ultimately, of observable relations between events, then this analysis will be a disappointment. But then it was never plausible that *causation* was likely to be analysed in terms of *more* familiar concepts. Causation is one of the very first concepts we acquire, certainly before we acquire any language, and arguably the concept is innate. As for the aim of finding an analysis in terms acceptable to a die-hard conceptual empiricist, that aim is as forlorn as the more general and moribund project of conceptual empiricism. On the other hand, if one's aim is to relate causation to metaphysically more fundamental features of the world, then the analysis provided does have something to offer.

7 Conclusion—causation and explanation

The following three facts have wide currency among philosophers, in the sense that each individually is held to plausible by many philosophers:

- (i) Explanations are supposed, optimally at least, to provide sufficient conditions for their explananda. An explanation is often supposed to show why we could (armed with the explanation) expect the explanan-

dum, which it could hardly do if it failed to provide sufficient conditions. Hempel's D-N model requires a non-elliptical explanation to be such that the explanandum is deducible from the explanans—hence the explanans is sufficient for the explanandum. An explanation may provide *conditionally* sufficient conditions for its explanandum, i.e. conditions that are sufficient given certain background conditions. An elliptical D-N explanation is of this sort.

(ii) Causes are counterfactually necessary conditions of their effects; they are typically not sufficient conditions. This is the thrust of the Hume-Lewis counterfactual account of causation, according to which causation is understood in terms of counterfactual dependence: had the cause not occurred, the effect would not have occurred. Something can fulfil this condition and so be a cause without being such that it is sufficient for the relevant effect.

(iii) To identify a cause of an event is to provide an explanation of it. Even if causal explanations are not the only explanations there are, they do form one prominent species of explanation.

There is a tension among these claims. If causal explanations cite causes as explanations of effects, as (iii) tells us, then, according to (i), causes should then be at least conditionally sufficient for their effects. But according to (ii) causes are conditionally necessary, not conditionally sufficient for their effects.

One might suspect (i), holding that the theory of explanation is not in very good shape. After all, has not Hempel's account been roundly refuted? Yes, but note that the principal and most powerful objections to Hempel's model are to the claim that sufficient conditions for *e* provide an explanation of *e*. Achinstein's (1983) poisoning case, for example, describes conditions that suffice for Jones's death (his ingesting a pound of arsenic) but which do not explain his death (because he was in fact killed by a bus). We are obliged by such examples to conclude that sufficient conditions do not always provide explanations, not that explanations do not provide sufficient conditions; to that extent (i) remains intact. (This is a matter to which I shall return.)

In this paper I have suggested that it is (ii) that is at fault. Causes are in a certain sense sufficient for their effects. For much of the discussion I supposed that a kind of conditional sufficiency, subjunctive sufficiency, provided the answer. But as we have seen, subjunctive sufficiency is in fact, for several reasons, only a proxy for the relation of a stimulus of a disposition to its manifestation. While in complex cases we may be able to explain what that relation is, in basic cases we cannot. But to acknowledge that inability to explain the basic manifestation relation is not an admission of failure. For given the initial commitment to an ontology of properties and essentially dispositional, one might expect the relation of disposition to manifestation to be basic also. What we have learned is, in the first place, that there is indeed a close relationship between dispositions and causation—we do not need to take a detour via laws and conditionals in order to say what causation is.

Along the way we have learned that insofar as there is some relationship between causes and conditionals, the relationship is closer to subjunctive conditionals than to counterfactual ones. (I have suggested that the perceived relationship between

causation and counterfactual dependence arises from conflating responsibility for an occurrence and causing it.)

As a consequence of the focus on subjunctive sufficiency, we have learned that the entirely natural distinction between cause and condition is an entirely legitimate one, metaphysically speaking. That in turn leads me to suggest that in some cases several events may be part of one collective cause (each part is not itself the cause but is part of the cause).

Thus this sketch of a dispositional approach to the metaphysics of causation reveals much about the nature of causation that has been hidden by ever more desperate attempts to capture it within a framework of counterfactual dependence.

References

- Achinstein, P. 1983. *The Nature of Explanation*. Oxford: Oxford University Press.
- Armstrong, D. M. 1983. *What is a Law of Nature?* Cambridge: Cambridge University Press.
- Bird, A. 1998. Dispositions and antidotes. *Philosophical Quarterly* 48: 227–34.
- Bird, A. 2005. The dispositionalist conception of laws. *Foundations of Science* 10: 353–70.
- Broadbent, A. 2007a. *A Reverse Counterfactual Analysis of Causation*. Ph. D. thesis, University of Cambridge.
- Broadbent, A. 2007b. Reversing the counterfactual analysis of causation. *International Journal of Philosophical Studies* 15: 169–89.
- Broadbent, A. 2008. The difference between cause and condition. *Proceedings of the Aristotelian Society*.
- Dretske, F. 1977. Laws of nature. *Philosophy of Science* 44: 248–68.
- Ellis, B. 2001. *Scientific Essentialism*. Cambridge: Cambridge University Press.
- Gundersen, L. B. 2004. Outline of a new semantics for counterfactuals. *Pacific Philosophical Quarterly* 85(1): 1–20.
- Hájek, A. 2007. Most counterfactuals are false. *Unpublished article*.
- Hall, N. 2004. Two concepts of causation. In J. Collins, N. Hall, and L. A. Paul (Eds.), *Causation and Counterfactuals*, pp. 181–276. Cambridge, MA: MIT Press.
- Johnston, M. 1992. How to speak of the colors. *Philosophical Studies* 68: 221–63.
- Lewis, D. K. 1973. *Counterfactuals*. Oxford: Blackwell.
- Lewis, D. K. 1986. *Philosophical Papers*, Volume II. New York, NY: Oxford University Press.

- Martin, C. B. 1994. Dispositions and conditionals. *Philosophical Quarterly* 44: 1–8.
- Mumford, S. 2004. *Laws in Nature*. London: Routledge.
- Swoyer, C. 1982. The nature of natural law. *Australasian Journal of Philosophy* 60: 203–23.
- Tooley, M. 1977. The nature of laws. *Canadian Journal of Philosophy* 7: 667–98.