

*forthcoming in*

Anne Sophie Meincke (ed.) *Dispositionalism.  
Perspectives from Metaphysics and the Philosophy of Science*

## A DISPOSITIONAL ACCOUNT OF CAUSATION, WITH SOME REMARKS ON THE ONTOLOGY OF DISPOSITIONS

Alexander Bird  
University of Bristol  
alexander.bird@bristol.ac.uk

### Abstract

I outline and consider the prospects for (and significance of) an account of causation in terms of dispositions. The basic idea is (roughly) that  $C$  causes  $E$  when  $C$  is the stimulus of some disposition of which  $E$  is the manifestation. 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. Finally I review the significance of this account of causation for a metaphysics of powers (potencies, essentially dispositional properties). I argue that it would be a mistake to see a successful dispositional account of causation (whether this or another account) as vindicating an ontology of powers.

**Keywords** dispositions, causation, ontology, conditionals

## 1 Introduction—dispositions and causes

What is the relationship between dispositions and causes? One approach to understanding dispositions analyses them in terms of causes. Lewis (1997) supplies the paradigm of this approach. The basic idea is:

(CD)  $o$  is disposed to manifest  $m$  in response to stimulus  $s$  iff  $o$  and  $s$  together cause  $m$

Lewis tweaks this to avoid problems arising from the possibility of finks. But the basic idea is clear and plausible enough. Some philosophers (e.g. Jacobs 2007; Bird 2010; Mumford and Anjum 2010, 2011), however, have sought to reverse the order of explanation, taking dispositionality to be more basic. Mumford and Anjum take this to provide significant ontological lessons. I disagree, and I shall return to that issue towards the end of this paper.

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 that in effect reverses (CD):

(SD)  $c$  causes  $e$  iff for some object  $o$  and some disposition  $D$ ,  $o$  possesses  $D$  and  $c$  is its stimulus and  $e$  is the manifestation of this instance of  $D$ .

The strategy I have for assessing (SD) is as follows.

- I shall first assume that there is a simple relationship between dispositions and conditionals, (CA). This will allow me to investigate a proposition, (SD–CA)', that is a consequence of (SD) *plus* (CA). (SD–CA)' identifies causation with a certain kind of conditional, which I call *subjunctive sufficiency*. I show that the subjunctive sufficiency view has significant merits.
- I then look at the weaknesses of the subjunctive sufficiency view, i.e. of (SD–CA)'. I argue that the weaknesses in (SD–CA)' are not due to (SD) but to (CA). I conclude that (SD) is a plausible account of causation: it has the strengths of (SD–CA)' but can avoid its weaknesses.

This strategy should be seen as analogous to the investigation of a scientific theory (such as the kinetic theory of heat) by making certain simplifying assumptions (e.g. that molecules are point particles with no attractive forces between them). We first see that the theory plus simplifying assumptions deliver fairly accurate results in important cases, delivering a good degree of confirmation. We see also that the theory delivers inaccurate results in other cases (e.g. when gases are under high pressure). We then argue that the imperfections are due to the strict falsity of the simplifying assumptions, and that the inaccuracies diminish when the simplifying assumptions are replaced by more realistic ones. This delivers further confirmation. In our case using the simplifying assumption (CA) has two 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, 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 of causation.

## 2 (SD) plus the simplifying assumption—the conditional analysis of dispositions

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

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

(CA) is false. But it isn't far from the truth. The standard counterexamples, finks, masks/antidotes, and mimics, require fairly unusual setups (Johnston 1992, Martin 1994, Bird 1998). (CA) is for many purposes a good approximation to the truth. As I explain later the standard counterexamples mentioned do not bear on the questions at issue in this paper. Nonetheless we will consider other cases where (CA) fails and will see what implications those failures have for the current project.

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

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

The existential quantification 'for some *o*' 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.) 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 conditional statements 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*. Subjunctive sufficiency 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 (after an initial clarification regarding the semantics of ' $\Box \rightarrow$ ') 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 objections can be explained away as arising from the counterexamples to (CA). To that extent (SD) remains intact. I shall conclude by considering the extent to which (SD) provides an illuminating account of causation.

### 3 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).<sup>1</sup>

There are indeed good reasons to reject centering. In the current context, one is the fact that (C) and (CA) together entail that every two actual events are dispositionally related. One might regard that as further proof that (CA) is defective. But

<sup>1</sup>Rejecting 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, p. 12-13.

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 a neighbouring laboratory violently rocks the building, causing the glass to shatter a fraction of a second 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.

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 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 cause of the shattering that is not the striking.

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’ and ‘if I were to strike it, it would shatter.’ According to that worry, 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 (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,<sup>2</sup> 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  $c$ .

So:

- (a)  $\wedge$  (C) entails (b);
- (b)  $\wedge$  (CA) entails (c).

In scenario III we have an entirely unobjectionable 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 III we thus accept the conditional but deny the disposition: 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, p. 46–8), after remarking that most people would be puzzled by, rather than assent to, counterfactuals joining unrelated actual facts, such as ‘If Canberra were the capital of Australia then the moon would have large craters’, goes on to point out that matters 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.’<sup>3</sup>

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 (1981), 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 (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 relevant to the truth of subjunctive conditionals, although I shall also consider the benefits of adding the restriction to normal worlds.

<sup>2</sup>At least if we assume:  $x$  is fragile  $\equiv x$  is disposed to shatter when struck.

<sup>3</sup>Hájek also discusses cases that turn on the incompatibility of ‘If  $x$  had occurred,  $y$  might not have happened’ and ‘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.

## 4 Causes and conditions

Lewis, along with many other philosophers who take counterfactual necessity as the mark of causation, draws no metaphysical 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.<sup>4</sup> 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. It was the flicking of the switch that 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.<sup>5</sup> Those who think that causes are counterfactually necessary conditions must regard all of these events as causes.

An advantage of the subjective sufficiency account (SD-CA)', and of the dispositional account (SD) it approximates, is that, unlike the counterfactual necessity view, the view respects and explains the natural cause-condition distinction.

Lewis (1986, p. 162) holds the cause-condition distinction to be merely pragmatic—all the events mentioned above 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.

Likewise Hall (2004, p. 228) sermonizes (his term),

Suppose that my favorite analysis counts the big bang as among the causes of today's snowfall... How easy it is to refute me, by observing 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 on 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, if possible, ought to be accommodated by a satisfactory theory. But here those reactions are dismissed as infected by pragmatics.

<sup>4</sup>The 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:  $\neg c \square \rightarrow \neg c$  (see also Broadbent 2007b). This is a sufficiency account of causation and so, like mine, generates the cause-condition distinction.

<sup>5</sup>Those 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.

The resort to pragmatics is an insufficient response. For a start, as Broadbent (2008) 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 (2004, p. 147) points out.<sup>6</sup> Moreover, salience and pragmatic concern may well focus one's interest on a necessary condition without thereby elevating it 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.

For a contrasting account of the cause-condition distinction, let us now turn to (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  $c' \square \rightarrow e$  is also false. The presence of oxygen is not a cause of the lighting. Likewise, *mutatis mutandis*, for the other cases.

(SD-CA)' is being employed as an approximation to the dispositional account (SD). We can also see directly that the latter respects the cause-condition 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.).

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, without a satisfactory account of the pragmatics at work, it lacks an explanation of why we make that error. The dispositional view of causation, however, because it respects our intuitive distinction between cause and condition, has a significant advantage.

## 5 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 setup 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,

<sup>6</sup>Menzies uses this to motivate a contextual account that retains difference-making.

[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, but I shall argue the opposite—our pre-theoretic judgments favour what I call collective causes: causes whose parts are not individually causes. I suggest that philosophers may be inclined to think that the fuel leak and the spark are each causes 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’. No, what caused the fire was the *co-occurrence* of the fuel leak and the spark. 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 two 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 (2005), 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.<sup>7</sup>

Again, the fact that the dispositional account, both as (SD–CA)’ and as (SD), explains the pre-theoretic verdicts, but the counterfactual necessity view does not, favours the former. (SD–CA)’ tells us:

[there is a fuel leak  $\square \rightarrow$  there is a fire] is false;  
[there is a spark  $\square \rightarrow$  there is a fire] is false; but  
[there is a fuel leak and there is a spark  $\square \rightarrow$  there is a fire] is true.

And likewise,

[the left engine fails  $\square \rightarrow$  the aeroplane crashes] is false;  
[the right engine fails  $\square \rightarrow$  the aeroplane crashes] is false; but  
[the left engine fails and the right engine fails  $\square \rightarrow$  the aeroplane crashes]  
is true, etc.

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. Thinking about the fire directly in dispositional terms, 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 was not disposed to bring about a fire in response to either of those events individually. The plane was disposed to crash if both its engines failed, but not disposed to crash if just one failed.

---

<sup>7</sup>MacBride is concerned to show that causation is a multigrade relation.

## 6 Counterfactual dependence?

Ned Hall (2004) 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 (2004, p. 253) 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, and so on. [Emphases in original.]

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 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—we are rather more interested in that 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—although strictly false—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. Mary 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 Mary 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 since they correlate it is not surprising that responsibility, causation, and necessary conditions are held to be intimately related. 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.

## 7 Problems and 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 coun-

terexample 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 the dispositional view can use it also. However, there are doubts as to whether the rejection of backtracking counterfactuals is legitimate (especially as Lewis's own small miracles employ them). The dispositional account has an independent answer that does not appeal to the denial of backtracking. 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 Canberra is the capital of Australia].

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.)

It may be noted however that none of the objections considered so far relates to the standard counterexamples to (CA), viz. finks, antidotes/masks, and mimics. We can see why not. Fink and antidote cases occur when there is a disposition that receives its characteristic stimulus but because the disposition is removed or interfered with it does not produce its manifestation. So those are cases where an event (the non-actual manifestation) does not occur. What we need to test an account of causation are cases of two *actual* events—the question is whether the account correctly classifies them as causally related or not. More relevant are cases of mimics and the finking production of a disposition. In the former case a cause brings about an effect in a way that mimics a disposition that is not present. For example, a robust iron pot is attached to a powerful bomb and a sensitive detonator. Strike the pot, the

bomb explodes, and the pot shatters—mimicking fragility. Such a case is not any problem for (SD–CA)' or (SD) in particular. In another kind of case one event both brings a disposition into existence (which was not present when the event occurred) and also is the stimulus of the disposition, so bringing about its manifestation. (CA) wrongly suggests that the disposition in question was present when the initiating event occurred. In Martin's (1994) example, an electro-fink device is attached to a dead wire, but makes it live if a conductor touches the wire—the conductor thus experiences a current. So the wire behaves as if it were live at the moment of touching even though it is not. Although the wire is not live, we can say that the overall setup (wire plus electro-fink) is disposed to conduct a current to a conductor that touches it. And so we are able to get the result that (SD) predicts.

## 8 Ontological considerations (I)—complex dispositions and the non-transitivity of causation

Dispositions are often made of combinations of simpler dispositions. My alarm clock is disposed to ring when the hour hand reaches a certain point. Clearly the latter event is the stimulus to the first of a series of dispositions, the last of which manifests itself in the ringing of the clock. Do we always have a disposition when we can find a sequence of dispositions like this? It is not clear that we do. A strong gust of wind causes snow to dislodge from a precipice. That initiates an avalanche. The avalanche flattens a tree in its path. In each case we may identify a disposition: the precariously balanced snow drift was disposed to fall in response to a gust of wind; the accumulated snow in the valley was disposed to become an avalanche in response to a disturbance, such as the falling snow drift. The avalanche was disposed to flatten trees in its path. Should we say that something (the valley?) was disposed to flatten the tree in response to the gust of wind? That does seem to stretch things, to say the least. Hitherto I have said nothing on the topic, except to accept that we can be entirely liberal in ascribing dispositions. But that assumption has not played any role in the arguments above. Here I want to consider the advantages of being somewhat more restrictive in our ascriptions of dispositions, restrictive in a way that is in tune with our disinclination to ascribe an encompassing complex disposition in the valley case.

Note first that our disinclination in that case mirrors our disinclination to ascribe a causal relation. We are *not* inclined to say that the gust of wind caused the tree to be flattened. Now, once again, that could be ascribed to pragmatics. But the explanation I suggest here is the relationship between causation and dispositions given by (SD). It does not matter that we do not absolutely reject the ascription of causation or of a disposition—both concepts are vague. What does matter is that the degree of disinclination is correlated, which confirms (SD).

The fact that we do not endorse immediately the causal connection suggests that our concept of causation is not such that we take it to be obviously transitive. Lewis, in defining causation as the ancestral of the counterfactual dependency relation, makes causation essentially transitive. That is despite the existence of counterexamples to the transitivity of causation. The dispositional account, if it is not entirely liberal with its ascription of dispositions, will also reject the transitivity of causation. Here is a characteristic example: Kvarn suffers an industrial accident in which his finger is severed. He rushes to the surgeon who skilfully reattaches it so that a year

later the finger is as healthy as it ever was. The accident causes Kvat to have surgery, the surgery causes the finger to be healthy at the later date, but the accident does not cause the finger to be healthy at that date.<sup>8</sup>

We can see that such examples substantiate the explanation of the failure of transitivity that I propose: the fact that dispositions do not always concatenate into more complex dispositions, although they sometimes do. So although Kvat is disposed to get surgery for a severed finger and the surgeon is disposed to repair a severed finger with skill such that it is healthy a year later, we do not think that Kvat has the disposition to have a healthy finger in response to that finger being severed. Nor would we attribute any such disposition to the sum of Kvat and the surgeon nor any other entity.

It is natural to ask then: when do dispositions concatenate to produce a complex disposition and when do they not? It is not possible to give a simple answer here. But here are some pointers:

- This is an analogue of the special composition question for objects. Not any set of objects compose to form a further object, but sometimes they do. But finding the correct principle or principles of composition is highly non-trivial. Note that the vagueness surrounding composition of dispositions is not a strong objection to the idea that they do sometimes concatenate but not always: the same issue arises for the composition of objects, but should not drive us to either unrestricted composition (every set of objects forms a composed object) nor to nihilism (no non-singleton non-empty set of objects ever forms a composed object).
- Easy cases of combination are where we can identify a whole with a function which is realised by the component dispositions. The functions may be natural or artificial. For example, the cardio-vascular system is disposed to transfer oxygen from the air we breathe to the muscles and other organs. That disposition is made up of the disposition of the lungs to dissolve oxygen from the air in the blood stream and the disposition of the heart to pump the blood to the muscles. On the other hand, some combinations of dispositions are clearly too accidental to form a further disposition, as in the cases discussed.
- One area to look to for further insights is the literature on *mechanisms*. Similar issues arise for the ontology of complex mechanisms as for complex objects and complex dispositions. A conjecture worth considering is the following: the complex dispositions are those combinations of dispositions that can be said to form mechanisms.

## 9 Ontological considerations (II)—powers

I have suggested that the ontology of dispositions is not entirely trivial. It is not as simple as saying that there is a disposition for every possible pair of stimulus and manifestation conditions. On the other hand, that fact tells us nothing about what

---

<sup>8</sup>This example is attributed by Hall (2000) to Kvat in connection with Kvat (1991). Some may find that such examples seem not to be examples of transitivity since the effects show no difference from the original state. But I think we get the same result even if we make changes to the final state. Imagine that the surgeon finds a small cyst and removes that while reattaching Kvat's finger. So the finger ends up in a healthier state than it was originally. Still, it is not the case that the accident caused the finger to be healthier than before.

dispositions in *are*. For all that has been said in the preceding section, dispositional properties may all or mostly be abundant properties—not genuine, ontic properties at all. Or they might be genuine properties, but not dispositional in nature or essence. Can we draw any conclusions from a dispositional account of causation regarding the nature of dispositions themselves?

Mumford and Anjum (2010, 2011) also offer a dispositional account of causation. They have a particular aim in their account, which is to vindicate an ontology of *powers*. The latter claims that properties (some at least) are dispositional in nature or essence. This contrasts with the view that no properties are by nature dispositional; rather a property might contingently have a dispositional character, but that character is conferred upon it by the contingent laws of nature—under different laws the same property might have had different effect and so a different dispositional character. Mumford and Anjum take the success of a dispositional account of causation to provide evidence in favour of the ontology of powers. So does the success of the current account (or indeed theirs) provide any such reason?

I do not think so. There is no clear inference from:

(A) All causation is the manifestation of some disposition.

to:

(B) Properties have a dispositional nature/essence.

Proposition (A) is consistent with the claim that the disposition that is associated with a given property depends on what laws of nature there are. The correctness (or otherwise) of this dispositional account of causation, (A), turns on facts about causal and dispositional relations in the actual world or worlds very much like it. Whether (B) is correct turns on the possibility (or not) of remote possible worlds with different laws and different dispositions associated with the same properties. So it is difficult to see how (A) could have any bearing on (B). The existence of powers is *one* way the world could come to have dispositionality and causation. Another way for there to be dispositionality and causation is for there to be contingent laws relating properties that are not powers (essentially categorical properties). The success of a dispositional account of causation (whether this one or that of Mumford and Anjum) is no better evidence for the former than for the latter. I suspect that any thought that such success supports the ontology of powers arises from a conflation of the concepts of ‘power’ and of ‘disposition’. Indeed Mumford and Anjum (2011, p. 4) explicitly say that they use these terms interchangeably. At best that is misleading and at worst begs the question. For on that equation a dispositional account of causation is by definition a powers account of causation. But that rules out by fiat the possibility that a non-powers ontology can equally well account for both causation and dispositionality. (I expand at length on these ontological issues in (Bird 2016).)

## 10 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 explanandum, 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 directed against *one* direction of the equivalence of explanation and sufficient conditions: the claim that sufficient conditions for *e* always 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. That leaves intact the view that to provide an explanation is to provide sufficient conditions in some form or other; so (i) survives.

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.

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).

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. Focussing on cases where dispositions and subjunctive sufficiency come apart allow us to avoid objections to the subjunctive sufficiency view of causation. Furthermore thinking about the ontology of dispositions shows us why counterexamples to the transitivity of causation occur. But they should not be expected to reveal anything about other fundamental questions of ontology, viz. whether properties are powers.

## References

- Achinstein, P. (1983). *The Nature of Explanation*. Oxford: Oxford University Press.
- Bird, A. (1998). Dispositions and antidotes. *Philosophical Quarterly* 48, 227–34.
- Bird, A. (2010). Causation and the manifestation of powers. In A. Marmodoro (Ed.), *The Metaphysics of Powers. Their grounding and their manifestations*. Abingdon: Routledge.
- Bird, A. (2016). Overpowering. (How the powers ontology has over-reached itself). *Mind* 125, 341–83.
- 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* 108, 355–64.
- Gundersen, L. B. (2004). Outline of a new semantics for counterfactuals. *Pacific Philosophical Quarterly* 85, 1–20.
- Hájek, A. (2007). Most counterfactuals are false. Unpublished article available at <http://philrsss.anu.edu.au/people-defaults/alanh/papers/MCF.pdf> last accessed 2015/10/19.
- Hall, N. (2000). Causation and the price of transitivity. *Journal of Philosophy* 97, 198–222.
- 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.
- Jacobs, J. D. (2007). *Causal Powers: A neo-Aristotelian metaphysic*. Ph. D. thesis, Indiana University.
- Johnston, M. (1992). How to speak of the colors. *Philosophical Studies* 68, 221–63.
- Kvart, I. (1991). Transitivity and the preemption of causal relevance. *Philosophical Studies* 64, 125–60.
- Lewis, D. K. (1986). *Philosophical Papers*, Volume II. New York, NY: Oxford University Press.
- Lewis, D. K. (1997). Finkish dispositions. *Philosophical Quarterly* 47, 143–58.
- MacBride, F. (2005). The particular–universal distinction: A dogma of metaphysics. *Mind* 114, 565–624.
- Martin, C. B. (1994). Dispositions and conditionals. *Philosophical Quarterly* 44, 1–8.
- Menzies, P. (2004). Difference-making in context. In J. Collins, N. Hall, and L. Paul (Eds.), *Causation and Counterfactuals*. Cambridge MA: MIT Press.

Mumford, S. and R. L. Anjum (2010). A powerful theory of causation. In A. Marmodoro (Ed.), *The Metaphysics of Powers: their Groundings and their Manifestations*. London: Routledge.

Mumford, S. and R. L. Anjum (2011). *Getting Causes from Powers*. Oxford: Oxford University Press.

Nozick, R. (1981). *Philosophical Explanations*. Cambridge, MA: Harvard University Press.