

CAUSATION AND THE MANIFESTATION OF POWERS

Alexander Bird

1

It is widely agreed that many causal relations can be regarded as dependent upon causal relations that are in some way more basic. For example, knocking down the first domino in a row of one hundred dominoes will be the cause of the hundredth domino falling. But this causal relation exists in virtue of the knocking of the first domino causing the falling of the second domino, and so forth. In such a case, A causes B in virtue of there being intermediate events $I_1 \dots I_n$ such that A causes I_1 , I_1 causes I_2 , \dots , I_{n-1} causes I_n , and I_n causes B. Cases of this sort include my putting my foot on the brake causing the car to slow, the smoke from a fire causing the fire brigade to be alerted, and so forth. In other cases the more basic causal relations may not be intermediate (or at least it is controversial that they are). My seeing that it is raining may cause me to want to stay inside, and this causal relation depends upon more basic causal relations among various components of my brain. But it does not seem possible to analyze this in terms of my perception causing certain brain events, which cause other brain events, which eventually cause my desire. Rather it seems as if the principle causal relation, between perception and desire, is constituted, rather than mediated, by the more basic causal relations in the brain. The same is true of the operation of the dynamo causing the current to flow. Again there are not intermediate events, but rather the causal relation between them is constituted by the motion of the charged particles in the wires moving through a magnetic field, which causes an electric field, which causes the charges to move in the wire.

There are thus at least two kinds of complex causal relation: the chain kind and the constitution kind. If we wish to understand causation, we need to understand the basic causal relations, at least as found in the chain kind. That is, to understand what it is for A to cause B when the latter is a causal relation of the chain kind, requires understanding what it is for the intermediate, basic causal relations to hold. In the case of a complex causal relation of the constitution kind, it is may be that understanding what it is for A and B to be causally related does *not* require understanding what it is for the constituting causal relations to hold. For a complex causal relation of the constitution kind may be itself regarded as basic relative to its own level of explanation.

This kind of distinction is implicit in David Lewis's account of causation (Lewis 1973a), according to which there is a basic causal relation, counterfactual depen-

dence, and that the general causal relation is the ancestral of this basic one: A causes B when A and B are related by a chain of basic causal (counterfactual dependence) relations. However, Lewis's motivation for this approach does not arise from the sorts of consideration given above but from the fact that counterfactual dependence fails to be a general account of causation thanks to the problem of pre-emption. Another motivation is the conviction that causation is transitive. Lewis ensures this by making causation the ancestral of counterfactual dependence.

2

What is the motor or cement of the universe? In David Armstrong's metaphysics it is the laws of nature (Armstrong 1983). For Lewis (1973b), there is a sense in which nothing is—all there is at the bottom is the Humean mosaic. Out of the latter he constructs the laws of nature, thence counterfactuals, thence causation. A third approach regards potencies (powers, essentially dispositional properties) as providing the ontological foundation. These different views *could* all agree on the analysis of causation—for example by agreeing with Lewis's analysis in terms of counterfactuals; arguably they might also agree on the analysis of counterfactuals in terms of possible worlds and laws ('and laws' because laws play a central role in determining the proximity of possible worlds). The different metaphysical views would then disagree on the analysis of lawhood: second order relations of necessitation; systematizations of the Humean mosaic; regularities generated by the potencies. However, one might wonder whether a more direct approach would be possible, identifying causation more directly with activity of the underlying ontology.

Consider then the ontology of potencies. Such entities are properties that are dispositional (and essentially so). Thus they are disposed to yield a manifestation in response to some stimulus. The fragile vase is disposed to break when struck, the elastic piece of rubber stretches readily when pulled. While fragility and elasticity may or may not be parts of our basic ontology, these properties illustrate the relation of disposition to manifestation and to stimulus. Furthermore, these examples seem to indicate a causal relation: the striking of the vase caused it to break, the pulling on the rubber band caused it to stretch. So, one simple proposal for an account of causation is as follows:

(SD) A causes B when A is the stimulus of some disposition and B is the corresponding manifestation.

Let us call this the simple dispositional analysis of causation.

The simple dispositional analysis of causation has interesting features. Given the presence of the dispositions and the absence of any interferers (finks and antidotes), the stimulus *suffices* for the effect. This sufficiency of is a subjunctive kind (were the cause to occur, the effect would occur), and so has modal force, but less than full metaphysical necessitation. Nonetheless, this subjunctive sufficiency of causes for their effects does mean that (in this subjunctive way) causes necessitate their effects. The latter is a plausible idea—causes make their effects happen—but it is an idea that has been largely ignored in recent literature, which has concentrated

on the Humean idea developed by Lewis that rather than being sufficient for their effects, causes are (counterfactually) necessary for their effects.

The counterfactual approach to causation has counterintuitive consequences, e.g. that the Big Bang is the cause of everything, that my birth is the cause of my death, that my being fed as a child is a cause of this paper being written, and so forth. The simple dispositional account avoids such consequences, since the first event in each pair is not a stimulus to a disposition whose manifestation is the second event: I did not have the disposition to write the paper in response to the stimulus of being fed earlier in life. Thus while the counterfactual approach makes no distinction between cause and condition, regarding our natural inclination to make such a distinction as explicable in terms of a pragmatic difference, the dispositional account respects a genuine difference: when the smoke sets off the fire alarm, the smoke is the cause whereas the presence of the alarm is just the condition, which corresponds to the fact that the smoke is the stimulus to the disposition of the fire alarm to sound when smoke is detected.

3

The simple dispositional account thus has certain advantages, which I elaborate on elsewhere. I shall now indicate a problem with the account. Dispositions can be mimicked: I strike an iron pot; attached to the iron pot is a bomb, which explodes just as I strike, destroying the pot. It looks as if the pot is fragile, although it is not. For this example is not even required that the striking cause the detonation of the bomb; they may be entirely coincidental. Now consider that the same scenario might occur, even with a fragile glass vase. I strike the vase, it is fragile, it shatters. But the cause of the shattering is not my striking, it is the bomb. So it looks as if we can have a scenario where some x is disposed to manifest Mx in response to stimulus Sx , stimulus Sx occurs, manifestation Mx occurs, but Sx is not the cause of Mx .

One response to the above is to focus on the details of the processes involved in the manifestation of a disposition. In the case where a striking does cause a fragile vase to shatter, there is a chain of cause and effect between striking and shattering, involving a certain pattern of stress originating at the point of striking which leads to small fractures, then to larger ones, which eventually join up, resulting in the shattering. In such a case the disposition has an underlying physical basis; that physical basis is ready to undergo the process described, when is activated by the stimulus. In the case where the bomb causes the shattering, this process might start but is superseded by a distinct pattern of causation starting from the bomb. So one might distinguish the first case from the second by requiring that each stage in the process that is the activation and playing out of the underlying basis of the disposition is completed.

While that answer may be satisfactory as far as it goes, it is insufficient, since the same problem might arise with basic dispositions which have no complex causal basis. Such basic dispositions or powers are simple properties, not constituted by any complex of properties. As a result the response of the previous paragraph has no purchase. But, arguably, the same problem may arise. If two distinct basic dispo-

sitions may have the same manifestation (as seems plausible) then we might have a situation where both dispositions receive their stimuli, but the manifestation is the effect of only one of these. Yet we cannot distinguish between the two dispositions on the basis of one having a complex basis that is fulfilled and the other having a complex basis that is only partially fulfilled., since neither has a complex basis. The case is similar to that of trumping preemption, due to Jonathan Schaffer (2000: 165):

SPELL It is a law that the first spell cast on any given day is enacted at midnight. At 1200 Merlin casts a spell to turn the prince into a frog and at 1800 Morgana casts a spell also to turn the prince into a frog. At 2400 the prince becomes a frog.

As the law dictates, the midnight transformation of the prince is an effect of Merlin's spell, not Morgana's. Schaffer specifies that the spell works directly, without intermediary events. Hence one cannot distinguish between Merlin's spell and Morgana's on the basis specified in the preceding paragraph. Schaffer shows that such examples reveal inadequacies in the counterfactual approach to causation, since it is false that had Merlin not cast his spell, the prince would not have become a frog (since Morgana's spell would have done the job). The problem for the dispositional account is different. Both spells have the power to transform the prince; both are enacted by the wizards; the manifestation appropriate to both occurs. Yet the manifestation (the transformation) is the effect of only one.

One might hold that SPELL is also a counterexample to the dispositional account, since in Morgana's case we have a manifested disposition without causation. To respond to this problem precisely it will help to get clearer about what the disposition and stimulus are. One possibility is that the disposition is a power belonging to the sorcerer, so each of Merlin and Morgana has the power to turn the prince into the frog, the stimulus of which is the uttering of the spell. In this case we can say that Morgana's power was *not* manifested; the appropriate event occurs, but not as a manifestation of *her* magical power. The correct conclusion to draw is that a disposition can receive its stimulus and an appropriate event of the manifestation-type occur, without that event being a manifestation of that disposition-token. (In effect we have a combination of an antidote (mask) to the disposition plus a mimic, as in the case of the fragile vase with bomb attached, considered earlier in this section.) Thus Morgana's power is no counterexample to (SD) or (SD') so long as we interpret those to mean that the effect event, B, is in fact a manifestation of *that* disposition, not merely an event of the manifestation-type.

Alternatively, the disposition in question is a disposition of the prince to turn into a frog, in response to the spell being uttered. In this case the changing of the prince into a frog is indeed a manifestation of that disposition. But it is not a manifestation in response to Morgana's utterance, but rather in response to Merlin's utterance. So (SD) and (SD') need further to be understood such that B is a manifestation-token of that disposition *in response to A*.

What SPELL and examples like it *do* show is the following. There is a difference between in fact being the manifestation-token of that disposition-token and merely being an event of the manifestation-type in terms of counterfactual or subjunctive conditionals. Likewise there is a difference between a disposition being manifested

in response to this stimulus-token rather than that stimulus-token, so that only one really simulates the disposition. The discussion of SPELL shows that SPELL shows that we cannot account for such differences in terms of counterfactual or subjunctive conditionals. The conclusion we must draw, in my view, is that the relation 'Mx is the manifestation of disposition Dx in response to stimulus Sx' is ontologically basic and is not reducible.

4

Now let us return to the idea that there are basic causal relations and complex ones. Consider:

DOMINOES A row of one hundred dominoes has been set up so that the knocking of the first domino causes the second domino to fall, causing the third to fall, . . . , causing the hundredth to fall.

The knocking of the first domino caused the falling of the hundredth. Is there a disposition somewhere such that the knocking of the first is the stimulus and the falling of the hundredth is the manifestation? Arguably the row of dominoes taken as a whole has such a disposition. But one might also think that there are cases of complex causation where it is less plausible to attribute a single disposition:

SPILLED TEA My elbow knocks over my cup of tea; the tea runs between the cracks in the floorboards; it shorts the lighting circuit; that trips the fuse; the lights go out in the kitchen.

So, it might be claimed, my moving my elbow caused the lights to go out in the kitchen; but one might be disinclined to say that there was an disposition for the lights to go out in response to my moving my elbow. Such a case is a counterexample to the simple dispositional analysis of causation.

It seems at least defensible to reject the causal claim here, and to retain the simple dispositional analysis. Is it really that clear that the movement of the elbow caused the lights to go out? Such a claim seems less intuitively assertible, that that the movement of the elbow caused the tea to spill, that the spilling of the tea caused tea to run between the floorboards, that the tea between the floorboards caused a short-circuit, and so forth. In 1879 Lt. John Rothery survived the battle of Isandlwana, at one point avoiding death by ducking to avoid a Zulu assegai. In 2008 his great-great-grandson writes an article about causation. A chain of causes therefore links the two events; nonetheless it seems distinctly odd to say that ducking an assegai in 1879 is the cause of the writing of the paper in 2008. The transitivity of causation is far from obvious, despite Ned Hall's assertions to the contrary (Hall 2004: 181). There are other well-known counterexamples in the literature (see McDermott 1995 for several cases). The general structure of many such counterexamples is that A causes B, where the function of B is precisely to negate the consequences of A-like events by causing C. Since C is incompatible with the continuation of A, A is generally not a plausible cause of C. For example:

WASTEBASKET FIRE A fire in a wastebasket causes the sprinkler system to operate, thus preventing (causing the absence of) a conflagration.¹

But the fire in the wastebasket is hardly a cause of the lack of conflagration. It strikes me therefore that there is ample room for simply denying the transitivity of causation. In which case we may hold that the movement of my elbow does not cause the lights to go out, and this parallels the lack of a dispositional connection between them. More generally one might suppose that our reluctance to ascribe causation in cases of chains of causes arises when one is reluctant to regard that chain as belonging to a single structure that could be regarded as the bearer of a disposition. In the case of the dominoes, they are set up precisely so that there will be a chain from the first domino to the last, and so is seen as a single system. Long separated historical events are not parts of one single system or structure. Events such as the wastebasket fire and the non-conflagration are most obviously not parts of the same system, since the first lowers the chance of the second. Note that in this case these events can be seen as components of overlapping systems: the sprinkler system has the wastebasket fire as its trigger; it also has the non-conflagration as its effect.

David Lewis (1973a) builds transitivity into his account of causation: the causal relation is the ancestral of the counterfactual dependency relation. One of the principal attractions of transitivity for Lewis is the fact that it allows him to escape certain pre-emption counterexamples to regarding causation as identical with counterfactual dependence. As a reason for favouring transitivity, this seems *ad hoc*, although Lewis does later give a defence (Lewis 2000). Nonetheless, if one does hold that it is clear that in SPILLED TEA the movement of the elbow caused the lights to go out, then one ought to make the parallel move to Lewis's, to regard causation as the stimulus-manifestation relation referred to in (SD), thus making our account of causation:

(SD') A causes B when there is a sequence of dispositions D_1, \dots, D_n such that for each D_i and D_{i+1} , the manifestation of the former is the stimulus to the latter, and A is the stimulus of D_1 and B is the manifestation of D_n .

Thus in SPILLED TEA we can say that the first event caused the last, since the position of the cup was such that it was disposed to fall if I moved by elbow, the gaps between the floorboards was such that they were disposed to let tea run between them if spilled, and so forth.

5

In the last section we saw two approaches to SPILLED TEA. Either one could deny that there is causation between the first and last event, holding that there is causation only between the neighbouring events in the sequence and that this is not transitive. Or one could hold that there is a causal relation between the first and last events, in addition those causal relations between the neighbouring events, and that causation is transitive. Either way, our focus must now be on the component

¹Nothing in this example depends on the final outcome being an absence.

or basic causal relations, those which the dispositional approach, whether as (SD) or as (SD'), holds are to be understood as the stimulus and manifestation of a disposition. For these (SD) gives the right answer. But does it provide any insight into causation? Furthermore, we saw as a result of considering SPELL that the stimulus-disposition-manifestation relation must itself be considered basic and metaphysically unanalysable, and so additional insight cannot be provided by further analysis of the right hand side of (SD).

One might argue that (SD) provides no insight into the nature of (at least basic) causal relations. We wanted to understand the causal relation but have just replaced it with the stimulus-disposition-manifestation relation. The latter is very much the same sort of thing as the causal relation, so we don't seem to have moved much further forward. And if anything the stimulus-disposition-manifestation relation is less well understood than the causal relation.

If what one wants from an analysis of causation is a reduction of the concept of causation in simpler, more familiar terms, then (SD) does not provide such a thing. Indeed, aspiring to such a thing seems rather a forlorn hope. What could be more basic or familiar than causation? Indeed, there is evidence that the concept of causation is innate, or that we are primed to acquire it very early on in life, within months at most. So even if there is a respectable philosophical task of conceptual analysis, causation does not look like the sort of concept that is going to get analysed in other terms. The only context in which the analysis of causation looks like a sensible project is one where one is laden with the baggage of empiricist assumptions about the nature of concepts: that all our concepts are ultimately derived from experience, and causation is not the sort of thing one can experience. Against that background it makes sense to find concepts which do derive directly from experience, and to try to analyse the concept of causation in terms of those. But if one does not share such empiricist presumptions, one has no reason to reject the thought that among our concepts, the concept of causation is about as basic as one can get.

The task of metaphysical analysis is a different matter. Here one might think that some categories of entity are complex entities composed of less complex entities. Thus, for example, Armstrong (1983) hold that laws of nature are not themselves basic entities, but are complexes consisting of first-order universals related by a certain second-order universal. The second-order universal ('N') is a basic entity, as also may be the universals that it relates.² As a conceptual analysis this would be a disaster, since 'N' is less familiar than 'law'. Furthermore, this doesn't seem much of an analysis since N and lawhood seem very similar sorts of thing. But such criticisms (parallel to those in the second paragraph in this section) miss the point of the exercise. If Armstrong is right we discover that laws are complex entities, and what they are complexes of, we discover what kinds of entity are required to be basic, including the crucial component 'N'. Furthermore, this provides, Armstrong argues, insight into other philosophical questions (such as the nature of properties and of causation, the problem of induction, and the Ravens Paradox).

Returning now to causation, we can see that similar responses are appropriate. While the analysis in terms of dispositions provides no conceptual reduction, it does

²I say 'may be' because Armstrong allows for some complex universals.

provide insight into the metaphysics of causation. We know (if the account is correct) what causation *is*—it is the stimulation and manifestation of a disposition. We must take some dispositions as basic natural properties, and this links to a more general metaphysics of powers or potencies, and thence to the metaphysics of laws. This much is non-trivial (it is inconsistent with most other accounts of causation). Furthermore, we gain insight into some of the philosophical problems of causation, for example, the distinction between cause and condition.³

References

- Armstrong, D. M. 1983. *What is a Law of Nature?* Cambridge: Cambridge University Press.
- Hall, N. 2004. Causation and the price of transitivity. In J. Collins, N. Hall, and L. A. Paul (Eds.), *Causation and Counterfactuals*, pp. 181–204. Cambridge, MA: MIT Press.
- Lewis, D. K. 1973a. Causation. *Journal of Philosophy* 70: 556–67.
- Lewis, D. K. 1973b. *Counterfactuals*. Oxford: Blackwell.
- Lewis, D. K. 2000. Causation as influence. *The Journal of Philosophy* 97 (4): 182–97.
- McDermott, M. 1995. Redundant causation. *The British Journal for the Philosophy of Science* 46(4): 523–544.
- Schaffer, J. 2000. Trumping preemption. *The Journal of Philosophy* 97(4): 165–181.

³I am very grateful to Michael Rota and audiences in Oxford and Buffalo for their comments.