

Event-causation: the counterfactual analysis

Jonathan Bennett

[From *Philosophical Perspectives* 1 (1987), pp. 367–386.]

1. Two distinctions

Philosophical theories of causation can be split two ways, into four groups.

First, there is the split between *event* and *fact* or *state of affairs* theories. Event theories give pride of place to causal statements like these:

- The explosion caused the fire.
- His sudden collapse was caused by the beating they gave him.

Fact theories give primacy rather to statements like these:

- The forest burned because an incendiary bomb was dropped into it.
- They beat him and that led to his suddenly collapsing.¹

The two are separated by the difference between perfect nominals such as ‘explosion’, ‘fight’, ‘divorce’, ‘quarrel’, ‘fire’, ‘collapse’, etc. on the one hand, and fully sentential clauses such as ‘They beat him’ and imperfect nominals like ‘his suddenly collapsing’ on the other.²

That split has to do with one’s selection of analysandum. The second split concerns choice of analysans: it is the split between *subsumption* and counterfactual analyses of causal statements. According to the subsumption analyses, ‘x caused y’ means roughly that x and y have properties that enable them to be subsumed under some causally true conditional—x under the antecedent, y under the consequent. Counterfactual analyses, on the other hand, say that ‘x caused y’ means something to the effect that if there hadn’t been x there wouldn’t have been y.

This paper will present a difficulty for the *counterfactual* theory of *event-causation*. If the difficulty is fatal, we have three options. **(1)** We can focus on event-causation statements, analysing these not as counterfactuals but rather as saying that ordered pairs of events fall under causal laws. **(2)** We can stay with counterfactuals but use them only to relate whole states of affairs, keeping the concept of an individual event out of it. **(3)** We can retreat still further, dropping both events and counterfactuals, and analysing causal statements as saying something about how pairs of facts, or pairs of states of affairs, can be subsumed under causal laws.

¹ Or ‘... caused him to collapse suddenly’. It would take too long to explain here why the noun-infinite form belongs with fact-causation rather than event-causation; but it does.

² For more on this, see Zeno Vendler, ‘Facts and Events’, in his *Linguistics in Philosophy* (Cornell University Press: Ithaca, N.Y., 1967), and my forthcoming book on the metaphysics of events. [*Events and their Names* (Indianapolis: Hackett, 1988).]

2. The counterfactual analysis

The counterfactual analysis of event-causation is one of the two offered by Hume. In its roughest, strongest form it says that

c caused e,

where c and e are individual events, means that e depended counterfactually on c, i.e. that

If c had not occurred, e would not have occurred.

This obviously has to be refined if it is to square with certain things we firmly believe about what causes what. The most obvious refinement is to equate 'c caused e' with the statement not that e depended counterfactually on c but rather that there is a series of events from c to e each member of which depended counterfactually on its immediate predecessor. (You may think that this is no refinement at all: if each member depends counterfactually on its immediate predecessor, doesn't it follow that the last depends counterfactually on the first? The answer is No. The relation expressed by counterfactual conditionals is well known not to be transitive.) Other refinements are also needed. For example, it can happen that c caused e but if c had not occurred some other event would have caused e; and the analysis must be modified so as to allow for that. Those matters are skilfully handled by the principal recent advocate of the counterfactual analysis, David Lewis, in his papers 'Causation' and 'Events', and I need not spend time on them.¹

There is another difficulty, however, which seems not to have been noticed before and which cannot be dealt with by minor repairs. It implies that no version of the counterfactual analysis of event-causation can be squared with our ordinary

ways of thinking: a defensible analysis along these lines would have to be radically revisionary—addressed not to the conceptual scheme we have, but to one that we could have—which is not what Hume and Lewis intended.

3. Essences of events

As Lewis's work brings out very clearly, any counterfactual about a particular event implies or presupposes something about the event's *essence*. If we are to counterfactualize about particulars we must be able to distinguish worlds at which a given event does not occur at all from ones at which it occurs but is somewhat unlike the way it actually was. Suppose that at noon precisely I wave my right hand, and someone makes the statement S:

If that hand-wave had not occurred, the auctioneer wouldn't have thought you were bidding.

Now, if I had waved my right hand a fraction faster than I actually did, or raised it an inch higher, the auctioneer would still have thought I was bidding; so if S is to come out true, those possible waves must count as the wave I actually did. That implies that my actual wave could have been a bit faster or higher than it was, which means that its actual speed and trajectory are not of its essence.

But suppose that at each of the nearest worlds where I don't wave my right hand at all at that moment I wave my left hand, and the auctioneer thinks I am bidding. Is S true in this case? You might think it isn't, because at those worlds my actual hand-wave doesn't occur and yet the auctioneer still thinks I am bidding. But *doesn't* my hand-wave occur at the left-wave worlds also? What is wrong with the idea that a single event which was in fact a right-handed wave could have been a left-handed one, so that the person who says

¹ David K. Lewis, 'Causation', *Journal of Philosophy* 70 (1973), 556–567, reprinted in his *Philosophical Papers*, vol. 2 (Oxford University Press: New York, 1986), in which 'Events' makes its first appearance.

'If that hand-wave had not occurred. . . ' is not pointing to worlds where at that moment I wave with my left rather than with my right hand? This raises the question of whether my right-handed wave was essentially right-handed. I don't want to answer it—just to illustrate its crucial relevance to counterfactuals about events.

There is not much literature on this. It comes under four headings.

(a) Some of it, including all that Davidson has said on the topic, is based directly on the fact that things like this are sometimes said:

Every Wednesday morning, the members of the squad vote on what practical joke to play on the corporal that day. Last Wednesday's joke was a hotfoot, but if pfc Jones had voted the other way it would have been a fake air-raid alarm.¹

If we take that as a datum, without processing or digesting it in any way, we shall conclude that a certain event which was a hotfoot could instead have been a fake air-raid alarm. This, in my opinion, amounts to refusing to take the notion of event-essence seriously; and I think that is how Davidson intended it—he was showing his contempt for counterfactuals about events, having been urged by Chisholm to say something about them.

It is pretty clear, anyway, that those standards assign wrong truth-values to many counterfactuals about events. In some circumstances we would want the counterfactual

If no member of the squad had had any matches, that hotfoot would not have occurred

to come out true. But it won't do so if the hotfoot could have been a fake air-raid alarm. And there is an easy way

of avoiding that unwanted result. We need only suppose that when the speaker says '. . . it would have been a fake air-raid alarm' he is using 'it' as a pronoun of laziness, to save himself from uttering 'last Wednesday's joke' again, and does not mean to refer to the same item as was first referred to by that phrase. Analogously, some say that in

John takes his family to a good restaurant every Christmas, and Henry does it every Thanksgiving,

the 'it' is clearly a pronoun of laziness. I shall say no more about this matter. It seems clear that if we are to get a useful view about the essences of events, it must have its roots in theory, and cannot be casually skimmed off the surface of our talk.

(b) Peter van Inwagen has likened the essences of events to the essences of substances.² We are inclined these days to believe that a substance's origin is essential to it: I could have become a farmer, but I could not have had parents other than my actual ones. Analogously, van Inwagen suggests, perhaps it is impossible that an event should have had a causal history different from the one that it actually had.

Van Inwagen's thesis clearly entails that if c caused e then if c had not occurred e would not have occurred; and this is also entailed by the simplest, boldest form of the counterfactual analysis of event-causation, though not by the analysis in its final, cautious version. Apart from that partial overlap, I have not sorted out the relations between the two. Anyway, I can't see how to put van Inwagen's idea to work in evaluating the counterfactual analysis, so I now set it aside without further discussion.

(c) Lawrence Lombard has argued that an event's time of occurrence is essential to it.³ Given that I waved my right

¹ See Donald Davidson. 'Eternal vs. Ephemeral Events', in his *Essays on Actions and Events* (Oxford: Clarendon Press, 1980), at pp. 197f.

² Peter van Inwagen, 'Ability and Responsibility', *Philosophical Review* 87 (1978), pp. 201–224. at pp. 208f.

³ Lawrence Lombard. *Events* (Routledge and Kegan Paul: London, 1986), pp. 206–216.

hand at noon precisely, I could have kept my hand in my pocket at that time and waved my right hand five seconds later, but that would necessarily have been a different wave. Lombard's one argument for this is unsound, I believe, but in any case we shall soon see that this view about the essences of events cannot be combined with the counterfactual analysis of event-causation.

(d) David Lewis, in his paper 'Events', says a lot about the essences of events. But rather than offering an independent theory about event essences, and then checking it against the counterfactual analysis of event-causation, he works in the opposite direction. He starts with our ordinary careful beliefs about what causes what, interprets them in accordance with the counterfactual analysis, and draws conclusions about what the essences of events must be like if we are not to be convicted of too much error in our views about what causes what. In the absence of any secure independent grounds for judgments about the essences of events, this modest procedure is acceptable, and may indeed be the best that can be managed.

4. An example and a result

Here is an example; it is mine, but it illustrates Lewis's procedure: I fall onto a lamp at midnight, knocking it to the floor so that it breaks. I stipulate that this happens in such a way as to make the statement

(i) My fall caused the lamp's destruction clearly true. I am not merely saying that my falling caused the lamp to be destroyed, or that the lamp was destroyed because I fell; those are *fact*-causation statements, and are quite irrelevant to my present topic. I am asserting that a certain particular fall caused a certain particular destruction, which is an *event*-causation statement. According to the counterfactual analysis, (i) is equivalent (near enough) to

(i) If my fall hadn't occurred, the lamp's destruction wouldn't have occurred either.

Now, suppose as is quite possible that (i) is true but that no conduct of mine could possibly have prevented the lamp from being destroyed within the next year: at all the physically possible worlds which are like ours up to just before midnight the lamp is destroyed within a year of that time. In that case, this counterfactual:

(ii) If my fall hadn't occurred the lamp would never have been destroyed

is clearly false. Now, if (i) is true while (ii) is false, it follows that the lamp's actual destruction—a particular event that I shall call D—has a richer essence than merely *being a destruction of that lamp*, or merely occurring when and where that lamp is destroyed. For if that were D's whole essence then D would exist at every world where the lamp is ever destroyed, including ones where I don't fall at midnight and the lamp is destroyed five years later. If some of those worlds are 'close' to the actual world, the statement that if my fall hadn't occurred D wouldn't have occurred comes out false, because at some of those close worlds my fall doesn't occur but D does occur five years later. And if that conditional is false, then—according to the counterfactual analysis—it is false after all that my fall caused the destruction of the lamp.

That gives us a negative result about the essences of events. There is more to the essence of any lamp's destruction than merely its occurring when and where that lamp is destroyed; there is more to the essence of any particular death than just its occurring when and where that person dies; and so on.

5. The asymmetry fact

Now, Lombard's theory about event essences has just this consequence: there is more to the essence of the destruction

of a lamp than its being a destruction of that lamp. It says that every event essentially occurs *when* it actually occurs, so that if D occurred at midnight any possible destruction of the lamp at any other time would have been a different event. So far, so good; but the theory has other consequences which cannot be safely combined with the counterfactual analysis of event-causation. Take a case where this is true:

There was heavy rain in April and electrical storms in the following two months; and in June the lightning took hold and started a forest fire. If it hadn't been for the heavy rain in April, the forest would have caught fire in May.

Add Lombard's thesis to that, and you get

If the April rain hadn't occurred the forest fire wouldn't have occurred.

Interpret that in terms of the counterfactual analysis and you get

The April rains caused the forest fire.

That is unacceptable. A good enough theory of events and of causation might give us reason to accept some things that seem intuitively to be false, but no theory should persuade us that delaying a forest's burning for a month (or indeed for a minute) is causing a forest fire.

But although you cannot cause a fire by *delaying* something's burning, you can cause a fire by *hastening* something's burning. When we judge that lightning caused this fire, we don't ask whether the forest would in any case have burned at some future time; and so the way is open for many events that we take to be causes of fires to be merely causes of something's burning earlier rather than later. Similarly, many causes of falls merely cause the thing

to fall earlier than it otherwise would have, and so on through causes of quarrels, reconciliations, thefts, slumps, recoveries, outbursts, landslides, floods, traffic jams, adjournments, and so on.¹

So perhaps we can combine the counterfactual analysis with half of Lombard's thesis, the half saying that if an event actually occurs at T then it essentially occurs no later than T. For then we can say, with respect to a hillside that slid at T and of which it is true that

If c had not occurred it would have slid later than T, that

If c had not occurred the (actual) landslide would not have occurred

which lets us infer that

c caused the landslide.

That looks about right: we treat as causes of landslides those events that speed up the land's sliding, but not those that delay it.

Actually, it is not quite as simple as that, because in special circumstances an event that stopped the forest from burning in May could also cause it to burn in June, and that event would be both a delayer of burning and a cause of the fire; similarly with the landslide. Here is another example: My disturbance in the church prevents John and Jane from marrying today; but it also affects one of the witnesses in a manner that leads him, months later, to bring John and Jane together again. In this case, my disturbance delays their marrying at T and is a cause of their subsequent marriage. A final example: A massage dislodges a blood clot that would have killed the patient within ten minutes, but also starts another chain of events that kills him two hours later.

¹ The point is especially clear in connection with causes of deaths. No event ever brings it about that someone dies whereas otherwise she wouldn't have died at all. But I keep deaths out of my illustrations because I find that lethal examples make philosophers suspect that I am trading unfairly on the special fact about deaths, namely that each of us undergoes exactly one of them.

Such complications are unimportant, however, and do not detract from my central thesis, which is that there is a strong asymmetry in this part of our conceptual scheme. Here is it in a nutshell:

You are informed that a movement of mine affected when a certain stone fell from the top of a wall; but for my movement, the stone would have fallen earlier than it actually did, or would have fallen later than it actually did; I'm not telling you which. I do tell you that fully informed observers of the scene agree about whether my movement caused the stone's fall, and I invite you to guess what their opinion is, i.e. to guess whether the movement did cause the fall. You have no basis for guessing.¹ Now consider: would it help you if I told you that but for my movement the stone would have fallen earlier than it actually did, or told you that but for my movement the stone would have fallen later than it actually did? Clearly that would help you. That reflects what I call the asymmetry fact about this part of our conceptual scheme.

6. Is it a conceptual fact?

I think that this is a conceptual fact. I contend that what we mean by '... is a cause of e' or '... causes e' is something of the form '... causes it to be the case that P_e at a certain time rather than later or never', where P_e is a temporally unsaturated proposition that is appropriately related to the event e. Thus, for example, to cause a fire is to cause it to be the case that the thing burns at a certain time rather than later or never, to cause a riot is to cause it to be the case that some people behave riotously at a certain time rather than later or never, and so on. If that is not right, then what is?

(1) The linguistic data might be explained in terms of

pragmatics rather than semantics. That is, it might be said that if a misunderstanding between two people delays their getting reconciled, the statement

The misunderstanding caused their reconciliation is, strictly speaking, true, but people are uncomfortable about assenting to it because it suggests something false, and this leads them to talk as though they could see it to be false.

It can happen that a statement is true in what it says and false in what it implies or suggests, and this can lead people to treat it as though it were false. And indeed this possibility can legitimately be used to defend a semantic thesis against apparent counterexamples. But we should never accept it in a particular case without asking how, why, the true statement comes to make the false suggestion; and in the present case there seems to be no decent answer.

Here is one try. 'Take a case where the cause delays the obtaining of a bad state of affairs: because of the nurse's therapy the patient did not have a stroke that morning though he did have one a month later. It is literally true that her therapy caused the patient's stroke, but we don't like saying this because it suggests something false, namely that *the nurse did something bad.*' But why should the statement suggest this if not because that is what it means? If the sentence 'The nurse's therapy caused the patient's stroke' means only that the therapy made a difference to when the patient suffered a stroke, why should that carry any suggestion at all that she did something bad, i.e. that she hastened his having a stroke rather than delaying it? Possible answer: 'Because hasteners are much more common than delayers. Given that the therapy made some difference to the time, it is statistically more likely to have brought it forward than to

¹ If you think you have. that is because you are nourishing your imagination on a one-sided diet of examples. arbitrarily picturing me as pushing the stone off the wall rather than protecting it from being pushed off.

have pushed it back.’ That would be an excellent answer if it were true. But it is blatantly false.

(2) Dropping the pragmatic approach, and conceding that the truth-value of ‘c caused e’ is pretty tightly tied to that of ‘c caused it to be the case that P_e at a certain time rather than later or never’, someone might suggest that what links them is not an immediate semantic connection but rather a fact about how the actual world is causally structured.¹ The suggestion is that there is a relation R such that: (a) it is a conceptual truth that c causes e only if $R(c,e)$, and (b) it is a contingent truth that $R(c,e)$ is seldom or never true when c delays the obtaining of P_e and usually or always true when it hastens it.

One could hardly accept this without being given some account of what R is, and I have no suggestions about that. Nor do I need to pursue the matter further. It is important for anyone wanting comprehensively to understand our concept of event-causation, but for my purposes here it makes no difference whether the asymmetry fact is purely conceptual or whether it is contingent. Either way, it will cause just as much trouble for the counterfactual analysis of event-causation, as I shall show in due course.

7. Trouble from the asymmetry fact

I shall pretend that the asymmetry fact is the fact that all hasteners and no delayers are causes. That is stronger than the truth, of course, but it will help to keep the discussion simple, and nothing in my argument will depend on that extra strength.

Let me be a little more exact about that. I shall first give my basic argument for the view that as long as our concept of event-causation is sensitive *in any degree* to the difference

between hasteners and delayers, the counterfactual analysis of event-causation is fatally flawed. Then I shall expound a possible way of escaping from its conclusion: it relies upon the notion of a ‘counterpart’, which will lead many to reject it out of hand; but I have no quarrel with counterparts, and will give two other reasons for disliking the suggested escape route. The more important of those reasons would fail if the asymmetry to which I have called attention were extremely weak—i.e. if there were only a mild tendency for hasteners to be causes of for delayers not to be. But, although the truth of the matter is not as strong as I shall (for expository purposes) pretend, it is nowhere near weak enough to undercut that final argument of mine. Here now is my basic argument against the counterfactual analysis.

The misunderstanding delayed their getting reconciled, so it did not cause their reconciliation. So, by the counterfactual analysis, it is not the case that

If the misunderstanding had not occurred, the reconciliation would not have occurred.

That means that the very same reconciliation occurred at some of the worlds where the misunderstanding did not occur. The only way of making sense of this is to suppose that at those worlds the reconciliation occurred earlier than it did at the actual world. In general, necessarily, *any event could have occurred earlier than it did occur*.

The intervention by the marriage counselor hastened their getting reconciled, so it caused their reconciliation. So, by the counterfactual analysis,

If the intervention had not occurred, the reconciliation would not have occurred.

But there are plenty of nearby worlds where the intervention didn’t occur and yet they did get reconciled weeks or months

¹ I owe this suggestion to Kit Fine, who has greatly helped me to clarify and focus my thinking in this paper.

later; so we must say that their actual reconciliation is not to be identified with any of those later reconciliations. In general, necessarily, *no event could have occurred later than it did occur*.

Putting these two results together, we get an incoherent position.

Start with an event e' at a world W' , and ask which events at other worlds can be identified with it. By the former of our results, e' can be identified with many events that occur earlier at other worlds—for example e at W . But now let us ask which events at worlds other than W we can identify e with. By the latter of our results, e cannot be identified with any events occurring later than it occurs at W —for example event e' at W' . And so, putting the two together: e' is e , but e is not e' .

To reinforce this, I shall work it out in terms of an example.

8. A concrete example

Consider two worlds, at each of which you and I are standing on top of a wall which is being demolished by a wrecker's ball; just in front of me there is a pebble, which I try to kick off the wall before the wrecker's ball gets to it. At world W I do kick the stone, with the result that it falls at T rather than a few seconds later; call that fall of the stone e . At world W' you push me so that my kick goes awry and the stone is undisturbed until the wrecker's ball knocks it off the wall a little after T ; call that fall of the stone e' . I add the further stipulation S , that the closest no-kick world to W is W' and the closest no-push world to W' is W . S doesn't follow from the rest of the description. but it is a consistent addition with it, and that is all we need.

Now, informed and competent speakers of English will agree that at W my kick causes e ; by the counterfactual

analysis that entails that if the kick had not occurred e would not have occurred; but by S the closest no-kick world is W' where e' occurs; so **e is not e'** . But competent speakers will also agree that at W' your push does not cause e' : by the counterfactual analysis it follows that it is false that if your push had not occurred e' would not have occurred; by S it follows that if your push had not occurred e' would still have occurred; the only possible candidate for the role of e' at W is e ; so **e' is e** . Thus we have a flat-out contradiction.

9. A couple of disclaimers

I haven't reached this result through illegitimately streamlining and strengthening the asymmetry fact. If there is *any* temporal asymmetry in this part of our causal thinking, there is trouble for the counterfactual analysis of event-causation. We are looking at a pair of worlds, W where e occurs at time t , and W' where e' occurs at time t' . And we are being forced—by the counterfactual analysis of event-causation, combined with the temporal asymmetry in this part of our conceptual scheme—to say that whether e is e' depends in part on whether the pair $\{t, t'\}$ exemplifies the later-than relation or rather the earlier-than relation. But this is absurd, because any pair that exemplifies one also exemplifies the other. And, as I said, the absurdity remains so long as there is any asymmetry, however hedged in by conditions, in this part of our conceptual scheme.

Furthermore, the trouble is just as bad if the asymmetry is not purely conceptual but reflects the fact that at the actual world hasteners tend to be causes while delayers tend not to be. Anyone who thinks there is such a contingent fact presumably takes it to be a deep, broad one—more like the fact that there is no action at a temporal distance than like the fact that I went swimming this morning—and so it will obtain not only at our world but also at worlds that are close

to ours, for example, worlds differing from ours only by a single kick or push and the fall of a pebble. That is all I need. In my kick-push example, for instance, the contradiction is reached just so long as the asymmetry fact obtains both at W and at W' ; it doesn't have to obtain at all worlds.

10. Privilege for the actual world

Perhaps there is a way out, however. Faced with a pair of event descriptions pertaining to two different worlds, we have been asking

What is the temporal relation between the events-at-worlds referred to by these descriptions? Is it earlier-than or later-than?

And that is an idiot question because if either answer is right then both are. But we might do better if there were some special feature that was always possessed by one but not the other member of our pair of event descriptions. Then we could replace the idiot question by something of the form

What temporal relation does the privileged one of these have to the other? Earlier-than or later-than?

and that might have either answer but cannot have both.

What could we mean here by 'privileged'? What is the special feature that is always possessed by just one member of the pair? The only remotely plausible answer is that the description involving the actual world is privileged. On that basis, then, the crucial question is something like this:

How does the event at the actual world relate to the event at the non-actual world? Is it earlier or later than it?

If we order our pairs in that way, by attending to which of them pertains to the actual world, we may seem to be on our way to the intuitively right answers. And this is just what the asymmetry fact implies if it is taken to be the fact that

No event could have occurred later than it *actually* did. Any event could have occurred earlier than it *actually* did.

But that is not the right way to express the asymmetry fact, and this proposed rescue of the counterfactual analysis is not acceptable.

Here is why. If we are to use the concept of event-causation at all, we should be able to use it not only **(a)** in describing what actually happens but also **(b)** in practical deliberation, where we don't yet know which of the worlds is actual, and **(c)** in counterfactuals about what would have caused what if things had gone differently in some respect. But neither the second nor third of these would go right if we accepted the proposed rescue of the counterfactual analysis, according to which we can't say whether c causes e at world W without knowing whether W is the actual world.

That this is true about **(b)** practical deliberation is perhaps too obvious to need to be spelled out in detail. As for **(c)** counterfactuals about event-causation: let us return to my kick-push example (near the start of section 8), and suppose that the actual world is W' at which you push me and my kick misses the stone. Intuitively it seems reasonable to suppose that at that world we could truly say that

If your push hadn't occurred, my kick would have caused a fall of the stone;

but by the counterfactual analysis, together with my stipulation S , this implies that

At world W it is the case that: if my kick hadn't occurred e wouldn't have occurred.

But the proposed rescue won't let us say this. It requires that none of the later falls at nearby worlds be identified with e , but now we are not assured of that because e does not occur at the actual world and the "it couldn't have occurred later" principle is now being applied only to actual events.

11. Counterpart theory to the rescue?

So far I have assumed that we are dealing with the identity relation between events at different worlds, but if instead we employ a *counterpart* relation we may after all be able to reconcile the asymmetry fact with the counterfactual analysis of event-causation. I shall explain this by first sketching Lewis's (and Leibniz's) counterpart theory for *substances*.¹

For simplicity's sake, let's narrow it down to people. The very same person cannot occur at more than one world, according to Lewis and Leibniz, but we can still divide counterfactuals about named people into true and false. Someone meditating on Nelson Rockefeller's handling of the Attica prison incident might say:

If Mario Cuomo had been in charge, there would have been no deaths or injuries;

and this has a chance of being non-vacuously true because it means something about possible men who qualify as *counterparts* of Mario Cuomo: they are sufficiently like him (in the right ways) for us to talk about *how they are* in the language of *how he might have been*. So the Cuomo counterfactual is true if, and only if, there are no deaths or injuries in the Attica prison incident at the closest world where a counterpart of Mario Cuomo is in charge.

For present purposes, the most important point to grasp is that identity is symmetrical whereas the counterpart relation need not be. In general, we expect the latter to go both ways: if possible person x is a counterpart of the actual Mario Cuomo, then Cuomo is probably a counterpart of x , so that some counterfactuals about x will get their truth-values from facts about Cuomo. But there is no necessity about this. Individual x at world W_x has as a counterpart y at world

W_y if and only if the two are alike in ways that have a certain kind of significance from the standpoint of W_x ; and y has x as a counterpart if and only if the two are alike in ways that have a certain kind of significance from the standpoint of W_y . The similarities are the same, whatever world you look at them from, but their kind or degree of significance may differ; and so the counterpart relation is not symmetrical.

This seems to open up a glittering way of escape from the difficulty which is my topic in this paper. For now we can say that at the world W where my kick causes the fall e of the stone, it is true that

If my kick had not occurred e would not have occurred,

because no event at the nearest no-kick world W' is a counterpart of e , and in particular e' is *not a counterpart of e* . But at W' where your push spoils my kick, we are still free to say that your push does not cause e' , implying that at the nearest no-push world some counterpart of e' does occur, to wit, *e is a counterpart of e'* . There is no contradiction in this, because being-a-counterpart-of is not a symmetrical relation.

12. An aside on modal continuants

David Lewis does not handle counterfactuals about events in the language of 'counterparts'. Rather than saying that the event we call the death of Socrates occurs only at the actual world though it has counterparts at other worlds, he treats the death of Socrates as what he has elsewhere called a 'modal continuant'—a class whose members belong to different worlds. An event, according to Lewis, is a class of spatiotemporal zones, no two existing at the same world. The class we designate by the phrase 'the death of Socrates'

¹ See David Lewis, 'Counterpart Theory and Quantified Modal Logic', originally published in 1968 and reprinted in his *Philosophical Papers*, vol. 1 (Oxford University Press: New York, 1983), pp. 26–39, with a 1983 Postscript, pp. 39–46.

has one member at the actual world, namely the place-time at which Socrates died.

Suppose that the informal English sentence ‘Socrates’ death could have been less than it actually was’ is true. According to counterpart theory it is true because

Some counterpart of Socrates’ death is less painful than his actual death was;

according to modal continuant theory, it is true because

Some member of Socrates’ death is less painful than its actual-world member was.

As that example suggests, there is a simple route from counterpart theory to the other: if you have been confining *x* to one world and handling counterfactuals about it in terms of its counterparts, you can change gear and instead identify *x* with the class whose members are it and all its erstwhile counterparts. I need not draw a map of the route for the journey the other way. It is clear enough that Lewis is right in saying that the theory of modal continuants ‘is an equivalent reformulation of counterpart theory’ (Ibid., p. 41.)

Although the two do the same work, they do it differently. Applying counterpart theory to substances, the referent of the name ‘Thomas Hobbes’ is perfectly determinate, but there is some indeterminacy about which possible things are counterparts of it, and this makes counterfactuals containing the name ‘Thomas Hobbes’ somewhat indeterminate. Applying modal continuant theory, there is indeterminacy about which item (which class of things) is picked out by ‘Thomas Hobbes’, but once the referent is fixed it is also determinate which worlds it exists at, and so counterfactuals in which the name occurs suffer from no indeterminacy from that source. Lewis is on record as objecting to such promiscuous indeterminacy of reference in the names of

substances (Ibid.), but it could be an advantage where events are concerned, for it is plausible to suppose that phrases like ‘Thomas Hobbes’s birth’ and ‘Thomas Hobbes’s reconciliation with Cromwell’ have referential slack that is not shared by the name ‘Thomas Hobbes’. Perhaps that is why Lewis chose to handle events in terms of modal continuants and substances in terms of counterparts.

Because the two theories are equivalent in power, the ‘nonsymmetry’ escape route that would be opened up by counterpart theory is also available to Lewis in the context of his treatment of events as modal continuants. But it seems more complicated there, and harder to grasp intuitively,¹ which is why I have chosen to present it in terms of counterpart theory.

13. Evaluating the proposed rescue

Someone who holds that counterpart theory is wrong, across the board, should think that I have already completed an adequate case against the counterfactual analysis of event-causation. Someone who rejects counterpart theory for substances, might nevertheless find it tolerable for events; but I cannot discuss that position, because I can think of no plausible reasons for it. What remains is the position of the person who is sympathetic to counterpart theory for particulars of every kind, including substances and events. Since that is my own position, I have a special reason for wanting to evaluate the proposed rescue of the counterfactual analysis from that standpoint.

One of Lewis’s reasons for counterpart theory is his ‘extreme’ realism about worlds: given his view of what sort of item a world is, it’s hard to see how a single substance could exist at more than one world. But even for those of us

¹ I am indebted to Kit Fine for a tutorial on how to do it.

who do not confidently agree with him about that, there are pulls towards counterpart theory as applied to substances. For me the strongest pull comes from the fact that answers to questions about the essences of substances seem not to fall sharply into the objectively true and objectively false, but rather to lie on a smooth scale from undeniable to intolerable, with much of the middle ground being sensitive to needs, interests, and contexts. If counterpart theory is true, that is just what one would expect; but if it is false—i.e. if questions of the form ‘Could x , which is not actually F , have been F ?’ have objectively right answers—it is puzzling that we should be so lost and adrift in our attempts to find out what the answers are.¹

This reason for favoring counterparts **(i)** rests on a general view about what is going on when in ordinary thought and talk we counterfactualize about particular substances, and **(ii)** owes nothing to the special needs of any philosophical theory in which counterfactuals are used or mentioned.

In sharp double contrast with that, the proposed invocation of counterparts in connection with events **(i)** cannot be motivated by reflection on our untutored untheoretical talk about how particular events might have been different, because we don’t engage in any such talk (or, if we do, we have no pretheoretic way of telling *when* we are talking in that way; consider the ‘practical joke’ example which I adapted from Davidson on page 3); and **(ii)** it owes everything to the fact that without it the counterfactual analysis of event-causation comes to grief. The difference between the two situations should, I submit, make us suspicious of the proposed treatment of events, inclining us to think it an ad

hoc rescue of a theory which should be allowed to fall.

Also: If event e occurs at world W at a certain time, and e' occurs at W' at the same time, it may be that each has the other as a counterpart. But for each such contemporaneous counterpart that e has, it has vastly many earlier counterparts of most of which it is not a counterpart. Thus, if the counterfactual analysis of event-causation is to square with the asymmetry fact, we need a drastically non-symmetrical counterpart relation—a relation R such that on the information that $R(x,y)$ the proposition that $R(y,x)$ is highly improbable.

I submit that this degree of non-symmetry makes the proposed rescue implausible. Our counterpart relation has to be induced through theory from the data concerning which counterfactuals we accept and which we reject, and that leaves room for the possibility that it is a non-symmetrical relation. But let us bear in mind what the role of this relation is supposed to be: it is a relation R such that

if $R(x,y)$ then the proposition that Fx can properly be expressed by saying that it could be the case that Fy .

It is not credible that a relation’s holding between x and y could have that effect—making us willing to say things about x in sentences that don’t name x —unless the relation approximated to identity, so to speak. So although strict symmetry should not be insisted upon, it is hard to believe that any counterpart relation should be as extravagantly non-symmetrical as the one we have been looking at.

So the counterpart relation cannot come to the rescue after all. The counterfactual theory of event identity cannot be squared with the asymmetry fact.²

¹ For some related remarks, see David Lewis’s *Counterfactuals* (Harvard University Press: Cambridge, Mass., 1973), pp. 38–43. Thomas McKay’s remarkable ‘Against the Constitutional Sufficiency Principle’, forthcoming in *Midwest Studies in Philosophy* 11 (1986) can be, though it is not by its author, used as a basis for a different though equally powerful case for counterpart theory.

² Stephen Leeds and Peter van Inwagen helped me greatly with this article.