

Causes, Effects, and Counterfactual Dependence

Kadri Vihvelin

There are at least two questions which any theory of causation should answer. It should tell us what it is for two events to be causally connected and it should tell us what makes it true that one event is the cause of the other, rather than the other way around. Hume offered an answer to both questions. He said that a particular event A and a particular event B are causally connected just in case events of the A kind are constantly conjoined with events of the B kind. And he said that A caused B just in case A occurred *before* B did. Thus, what makes it true that putting the log in the fire caused its burning is just the fact that all events similar to the first are followed by events similar to the second.

We know that Hume's account is oversimple, but many of us suspect that something like it must be the right story. That is, we've been convinced by Hume's arguments against the 'confused idea' that the causal relation is a necessary connection between particular events. (What would such a thing *be*? How would we ever know that it obtained?) And we accept Hume's claim that what makes it true that putting the log in the fire caused its burning is not (merely) a fact about *this* log and *this* fire, but is also some kind of general fact. We believe that the two particular events are causally connected just in case they satisfy some description which figures in some law.¹

But Hume's answer to the second question is not right. It's true that all the causes we've seen so far have occurred earlier than their effects. But it's not necessarily so. We can tell consistent stories in which it would be reasonable to believe that a cause occurs later than its effect -- stories about time travel, for instance.² So the difference between the fact that A caused B and the fact that B caused A is not *merely* the fact that A happened before B did. We need a theory of causation which does what Hume failed to do. We need an account which distinguishes causes from effects in a way that allows it to be a *contingent* fact that causes are temporally prior to their effects.

David Lewis offers such an account.³ He says that causation is to be analysed in terms of the counterfactual dependence of one event on another and he argues that event counterfactuals are contingently temporally asymmetric in the right kind of way. His account runs into problems with cases of causal

pre-emption, but so do the accounts which follow Hume more closely.⁴ I think that Lewis's theory of causation is the closest thing we have to an answer to the problem of distinguishing the cause-effect relation from the effect-cause relation. But it doesn't succeed.

I'll be arguing that Lewis's theory of causation fails because of the demands that it places on Lewis's theory of events. Some people think that causal relata are facts rather than events. That's not my complaint. Lewis is prepared to be quite generous in what counts as an event; on his view, events are just the entities that play the role of causal relata. If they turn out to be more factlike in structure than the battles, strolls, conversations, etc. we are used to thinking of as events, then that's fine with Lewis. But Lewis needs an account of events which is both *generous* enough to supply enough causal relata for all cases of causation and *restrictive* enough to ensure that effects don't cause their causes (given his account of what the causal relation is). I will argue that no theory of events can meet both demands.

I will use the failure of Lewis's account as a way of motivating my own counterfactual theory of causation. In doing so, I help myself to Lewis's possible worlds semantics for counterfactuals, but I do not assume the truth of his substantive theory of how we evaluate counterfactuals. In particular, I do not assume that the closest worlds are those at which most of the past is the same as ours. I will rely only on those beliefs about counterfactuals which should be treated as data by any adequate theory of counterfactuals.

Like Lewis's account, mine has the virtue that it is not an analytic truth that causes are temporally prior to their effects. I should make it clear that I do not, in this paper, undertake to solve what is sometimes called 'the problem of causal direction' - the problem of explaining the contingent feature of the world (entropy? the overdetermination of the past by the future?) in virtue of which it's true that causes are temporally prior to their effects. A complete defence of my theory of causation would require showing that this contingent fact, together with our similarity metric for counterfactuals, results in the right kind of temporal asymmetry of counterfactuals.⁵

What I do argue is that my account can draw the cause/effect distinction in all the cases where Lewis's account does, and that my account succeeds in places where Lewis's account fails. My claim is that Lewis's account works as well as it does because it's an approximation of my account. Lewis's account

get things more or less right in the cases we think of as standard cases of event-causation; my account works for other sorts of cases as well.

Let's begin by comparing Lewis's theory to mine.

Two Theories Of Causation

A very basic intuition about causation is that a cause is what makes the difference between the occurrence and nonoccurrence of its effect. By producing the cause, we produce the effect; by preventing the cause, we prevent the effect. As an account of causation, this is both anthropomorphic and circular, but the basic idea can be given a counterfactual formulation free of these defects: If the cause had occurred, the effect would have occurred and if the cause had not occurred, the effect would not have occurred.

Lewis's account *looks* as though it captures this insight. Lewis says that a particular event C caused another particular event E if it's true that:

(1) If C had not occurred, then E would not have occurred;

and

(2) If C had occurred, then E would have occurred.⁶

But this is misleading. According to the now widely accepted Lewis/Stalnaker semantics for counterfactuals, the second counterfactual is trivially true (since the actual world is closer to itself than any other world is to it, and C and E both occur at the actual world). So, where C and E are particular events which in fact occur, Lewis's account reduces the truth of 'C caused E' to (1). This captures the idea that by *preventing* the cause, we would prevent the effect, but it fails to capture the idea that by *producing* the cause, we would produce the effect.

If we combine Lewis's semantics for counterfactuals with Lewis's counterfactual analysis of causation, we lose one half of our intuitive conception of causation. Is the fault with the semantics or with Lewis's theory of causation?

You *could* blame the semantics. You could deny that the truth of P and Q entails that if P had been the case, Q would have been the case. You could insist that more is required for the truth of the counterfactual than the bare truth of P and Q.

But I think that this would be a mistake. Consider the following example. Smith drove the car and a highly freakish accident occurred. Jones, who knows nothing about the accident, and who thinks that Bloggs drove the car, says: 'If Smith had been driving, there would have been an accident.' Did Jones say something true or false? We can give two different answers. Jones said something true because, as it turns out, Smith *did* drive the car, and there *was* an accident. Jones said something false because the fact that Smith drove the car had nothing to do with the accident; the accident was caused by a freakish explosion in the gas tank, not by Smith's driving.

If we take Jones to be saying that if Smith had been driving, her driving *would have caused* an accident, then what Jones said is false. But while the causal interpretation of Jones's remark is the most natural one, it's not the only one. Jones might have been making the more modest claim that *if* Smith had been driving, there would have been an accident (whether because of Smith's driving or for some other reason). If so, then Jones said something true, since Smith was the driver, and there was an accident.

The moral we should draw is that the Lewis/Stalnaker semantics provides the right truth-conditions for counterfactuals with true antecedents *provided* that these counterfactuals are understood without causal implications. (Which, of course, is just what anyone seeking to give a counterfactual analysis of causation needs). But then Lewis's (2) is trivially true, and his analysis of causation fails to capture the idea that causes are nontrivially sufficient for their effects.

I think we can capture that idea, without abandoning our semantics, by replacing (2) with a different counterfactual. That is, where C and E are particular events which in fact occur, I think we should say that C caused E just in case it's true that (1) and *also* true that:

- (3) If neither C nor E had occurred, then (if C had occurred, E would have occurred).

Note that (3) is a counterfactual which embeds a counterfactual. That is, (3) says that at the closest worlds at which neither C nor E occur it's true (in those counterfactual circumstances) that *if* C had occurred, then E would also have occurred. This gives us a nontrivial way of understanding the

claim that C is counterfactually sufficient for E and thus a way of capturing the intuitive idea that a cause is a producer of its effect.

To see how this works, consider a pair of events: the accident and the death of the driver. Did the accident cause the driver's death? For Lewis, the answer turns simply upon whether or not it is true that:

- (1') If the accident had not occurred, the driver's death would not have occurred.

I think things are more complicated than that. Even if (1') is true, I don't think we can conclude that the accident caused the death. (1') tells us that preventing the accident would have *prevented* the driver's death. But it doesn't tell us that the accident *produced* the death. How do we tell that? Well, consider what things would have been like had the accident not occurred. That is, consider the closest worlds at which the accident doesn't happen. Now (1') tells us that at those worlds the driver's death does not occur. But I want to ask a further question: In those counterfactual circumstances (where neither the accident nor the death occur), would the death have occurred if the accident had? If in those circumstances - circumstances, remember, which are very similar to the actual circumstances - the accident *wouldn't* have produced the death, then I conclude that the accident didn't produce the death at the actual world either. In that case, I say that the accident did not cause the death. That is, I say that the accident caused the driver's death only if it's true that (1') and *also* true that:

- (3') If neither the accident nor the driver's death had occurred, then (if the accident had occurred, the driver's death would have occurred).

According to the logic of counterfactuals defended by Lewis⁷, (1') and (3') are both true just in case (4') is true:

- (4') If the accident had not occurred, then (the death would not have occurred *and* if the accident had occurred, then the death would have occurred).

This suggests a more concise way of stating my account. I propose that a particular (actually occurring) event C caused a particular (actually occurring) event E just in case (4) is true:

- (4) If C had not occurred, then (E would not have occurred *and* if C had occurred, then E would have occurred).

Lewis's account looks simpler than mine, and it gives the right result in a lot of cases. But there is a large class of cases in which his theory gives the wrong results, and I think that the reason things go wrong has to do with Lewis's way of distinguishing causes from their effects.

Lewis On Causes and Effects

Lewis takes causal connections between events to be a matter of the counterfactual dependence, as defined by (1), of one event upon another. The events we call 'effects' wouldn't have happened if it were not for the occurrence of the events we call 'cause'. But not vice versa. The events we call 'causes' would have occurred even if their effects had been somehow kept from happening. This asymmetry, Lewis holds, is all that distinguishes causes from effects. Now, as a matter of fact, so far as we know, it turns out that the events we call 'causes' always temporally precede those we call 'effects'. Future events counterfactually depend on past events, but not vice versa. But it ain't necessarily so. There are possible worlds where time travellers shoot dinosaurs in the Jurassic and in those worlds there are dino-deaths that would not have happened if the hunter had not got into the time machine 60 million years later. In those worlds there are past events that counterfactually depend upon the occurrence of future events, and in those worlds there is causation backwards in time. On Lewis's account, the direction of causation just is the direction of counterfactual dependence.

This is elegant. It allows us to distinguish causes from effects in a way that allows it to be a contingent fact that causes precede their effects. What we want to say, *roughly*, is that causes are temporally prior to their effects in virtue of the following contingent asymmetry of counterfactual dependence:

- (A1) If the past had been different, the future would be different; but if the future had been different, the past would not have been different.

But that's too rough. (A1) is false. It's easy to describe cases in which there would have to have been a change in the past, if the future had been different. Thus:

- (5) If Kennedy had served a second term as President, his assassination would not have occurred.
- (6) If Bush had been inaugurated in January 1993, then he would have won the election in 1992.

These are 'backtracking' counterfactuals, and they are true.

But notice something about the examples. The first is a fact-event backtracker. It says that if something had been true-- if something had been a fact-- then some event would not have occurred. The second is a fact-fact backtracker; it says that if something had been true, then something else would have been true.

Query: Are there any true backtracking *event-event* counterfactuals? That is, does it ever happen that there is a particular event e_2 and a particular event e_1 such that:

- (7) If e_2 had not occurred, then e_1 would not have occurred

where e_2 occurs later than e_1 ? Lewis is committed to the claim that this never happens. If he is right, then this is a striking difference between events and facts. And, however we ultimately *analyse* the difference between facts and events, our analysis had better preserve this striking fact. Maybe this explains why people tend to speak of causation as a relation between *events*. Certainly, it's a good reason for us to confine our counterfactual analysis to event-event causation. It would allow us to say that what makes it true that causes occur before their effects is that, *roughly*:

- (A2) The occurrence of later events counterfactually depends on the occurrence of earlier events, but the occurrence of earlier events never depends on the occurrence of later events.

But (A2) is still too rough.

There has been a terrible accident. A driver was speeding down the highway, a deer jumped in front of his car, he swerved wildly to avoid it, his car plunged over the embankment to the bottom of a gorge and exploded. The driver, horribly burned, died later, in the hospital. The accident caused the driver's death, and it is clearly true that if the accident had not happened, the death would not have occurred. Now (A2) says that it should be false that:

- (8) If the death had not occurred, the accident would not have occurred.

since the accident precedes the death. And this counterfactual does indeed seem false. We can imagine a world, very similar to our own, where the accident happens, but the gas tank doesn't rupture, or the driver is thrown clear, or the ambulance arrives more quickly... where something happens, after the accident, so that the death does not occur.

But, granting that this is so, we might insist that there is nevertheless a sense in which (A2) is false. After all, we want to say that if the death had not occurred, then *something* would have to have happened differently. That is, some prior event which actually occurred would have to have failed to occur; the gas tank would not have ruptured or the driver would have been thrown clear or been taken to the hospital sooner.... We don't want to say that if the death hadn't happened, everything which in fact happened would still have happened. Lewis would agree.⁸

(A2) is false because if the driver's death had not occurred, then some or other earlier event would not have occurred. But note that there is no *specific* event -- e.g. the rupture of the gas tank - that would have to have failed to occur.

Let's try again. What makes it true that causes are temporally prior to their effects is that event counterfactuals are temporally asymmetric in the following way:

Event Asymmetry Thesis: While there are many pairs of particular events e_1 and e_2 (where e_1 occurs earlier than e_2) such that if e_1 had not occurred, e_2 would not have occurred, there is no

pair of *particular* events e_1 and e_2 such that if e_2 had not occurred, then e_1 would not have occurred.

Or so Lewis must claim.⁹

Is the Event Asymmetry thesis true? I agree that it's hard to think of counter-examples if we restrict ourselves to the kinds of particular events we commonly talk about: accidents, deaths, births, assassinations. But that doesn't seem decisive. Why *can't* there be events-- *particular* events-- for which the Event Asymmetry thesis fails? What, anyway, is a particular event?

Events come in many shapes and sizes. The deer's darting across the road is an event. The driver's swerving is another, the explosion another still. The driver's death is a relatively small and localized event; it happens suddenly in a hospital bed. The accident is a bigger event. It includes other events - the swerve, the plunge, the impact. The accident caused the explosion, so they are different events, but together the accident and the explosion seem to add up to a bigger event, an event which led to the driver's being taken to the hospital.

Query: Is there an event which consists in the collective occurrence of *all the events* that happened from the moment the accident began to the instant just before death? Suppose that there is such an event-- the conjunction of all the events that actually lead up to the driver's death. Let's call this event 'Bob'. Then it's true that:

- (9) If the driver's death had not occurred, then Bob would not have occurred.

If Bob is an event, the Event Asymmetry thesis is false. Lewis's conclusion is that Bob isn't *really* an event. He claims that there can't be events which are as big and complicated as Bob. This, we might think, is a procrustean move; surely some events can be combinations of other events -- the accident is the combination of the swerve, the plunge, and the impact. What's so bad about Bob?

We'll look at Lewis's answer in a moment. First, let's look at a different kind of counter-example to the Event Asymmetry thesis.

Suppose Kennedy had won a second term as President. Imagine what Kennedy's second inauguration, in 1965, would have been like. It would surely have been a grand event -- Schlesinger speeches, Sandburg poetry, Camelot in

full bloom. Alas, this event did not occur. Kennedy's assassination prevented it. Question: Can we describe the *non-occurrence* of Kennedy's second inauguration as an event? Suppose there is such an event: the failure-of-the-second-inauguration-to-occur-in-1965. Let's call this event 'Oliver'. We could say things like:

- (10) If Kennedy's assassination had not occurred, Oliver would not have occurred.

Which might be true. More to the present point, if Oliver is an event, then it is, for sure, true that:

- (11) If Oliver had not occurred, Kennedy's assassination would not have occurred.

For Oliver's not occurring would consist in there being a second Kennedy inauguration, and if there were a second Kennedy inauguration, his assassination would have to have not occurred. But (11) is an event-event backtracker. So if Oliver is an event, the Event Asymmetry thesis is false.

Once again, Lewis has to deny that things like Oliver-- failures of other events to occur -- can be events. This may not seem as procrustean a move as banning Bob; Oliver looks like a philosophers' invention. On the other hand, in our ordinary talk about cause and effects, we do sometimes talk *as if* the failures of certain things to happen were causes: 'Kennedy's death occurred because the secret service failed to check the book depository'; 'Raoul's failure to wear a helmet caused his head injury.' If these aren't event-event causal claims then Lewis's analysis of causation can't handle them. But if the secret service's failure is an event, what *is* the difference between it and Oliver?

Lewis on Events

Lewis's theory of events goes like this:¹⁰ A particular event is the class of actual and possible spatiotemporal regions at which the event occurs. Events have accidental as well as essential properties. An event may occur differently without failing to occur. The departmental meeting was in fact a meeting of ten people which lasted an hour, but if fewer people had been

there, it --- the very same event --- would not have lasted as long. The essence of an event is what all the regions at which it occurs have in common.

Thus, the sinking of the Titanic is the class of spatiotemporal regions which have in common the fact that the Titanic sinks; the Titanic's sinking quickly is the class of spatiotemporal regions which have in common the fact that the Titanic sinks quickly. Events come in degrees of fragility, corresponding to how detailed the essence of the event is. The more detailed the essence, the smaller the number of regions at which the event occurs, and the more fragile the event. The Titanic's sinking quickly while men were heroic is a more fragile event than the Titanic's sinking quickly which is more fragile than the Titanic's sinking.

Lewis imposes certain restrictions on the classes of spatiotemporal regions which are eligible to be counted as events and on which events are eligible to stand in causal relations to each other. Some of these restrictions are motivated by Lewis's broadly Humean beliefs about causation, and I have no quarrel with these. He insists, for instance, that events are contingent entities and that only 'wholly distinct' entities are eligible to cause each other. The Titanic's sinking quickly and the Titanic's sinking are different events (for they have different causes and effects) but they are not 'wholly distinct' because the occurrence of the former implies the occurrence of the latter. (Necessarily, if the former occurs in a region, the latter also occurs in that region.)¹¹

But Lewis also needs further restrictions on what counts as an event; to ensure the truth of the Event Asymmetry thesis, he needs to rule out events like Bob and Oliver.

Lewis rules out events like Bob by stipulating that there are no events with essences which are so detailed that they include all or most of the intrinsic facts about a spatiotemporal region. His rationale is that such entities would be highly fragile; it would be too easy for them not to occur, making them unsuitable to be counted as causal relata.¹²

This seems arbitrary and ad hoc. Even if the events we *ordinarily* count as causes and effects are not extremely fragile, why rule out such events altogether? Extremely fragile events allow us to draw needed distinctions in certain kinds of cases. For instance, consider the classic overdetermination scenario in which someone dies after being struck simultaneously by two fatal bullets, one from Smith's gun and one from Jones's gun. On Lewis's account,

it's false that Smith's shooting caused the death, for if Smith's shooting had not occurred, Jones's bullet would still have done the job. And on his account, it's also false that Jones's shooting caused the death, for similar reasons. But if we don't ban fragile events, then we can distinguish between the very fragile two-bullet death which Smith's shooting caused (and which would not have occurred had his shooting not occurred) and the less fragile death which Smith did not cause (for it would have occurred even if his shooting had not occurred).

Lewis recognizes that there are contexts in which very fragile events allow us to draw relevant causal distinctions, but he says that there is 'no principled way' of distinguishing between the fragile events which we want to recognize and those fragile events which result in what he calls 'spurious causal dependence'.¹³ But none of the examples he gives are convincing. So far as I can see, the only grounds Lewis has for rejecting an event like Bob is that if Bob is an event, then Lewis's theory of causation is false.

We may have more sympathy with Lewis's argument against events like Oliver. The problem with events like Oliver, according to Lewis, is that they can occur in too many different ways, ways which intuitively have nothing in common. At the actual world, Oliver occurred as Johnson's inauguration. But there are possible worlds at which Oliver occurs as Barry Goldwater's inauguration, other worlds where it occurs as Henry Wallace's inauguration, and still other worlds where it occurs as Elvis Presley's inauguration. And so on.

Compare Oliver to the kind of event we ordinarily talk about. Yesterday's department meeting could have occurred in a number of different ways, but there are limits to how different it could have been. It might have been shorter or longer, it might have consisted of fewer or more of the department's members, it might have been more or less boring, more or less acrimonious. But the department's meeting would not have occurred at all had none of the department's members been present.

This suggests that what distinguishes philosophers' events like Oliver from *real* events like department meetings, conversations, and battles is that the different occurrences of Oliver are not sufficiently similar to each other. Lewis builds this intuition into his theory of events; he says that there are no events with essences which are disjunctive in 'overly varied' ways. Whereas the problem with Bob is that Bob is too fragile, the problem

with Oliver is that Oliver is too resilient -- it's too easy for Oliver to occur.¹⁴

We may wonder how the line is to be drawn. Lewis is quite generous in what he counts as occurrences of the same event; he says, for instance, that the shooting which was in fact done by Ned could have been done by Ted (where Ned and Ted are both members of the firing squad).¹⁵ He would admit, I think, that some of the different occurrences of the department's meeting are alike neither with respect to any individual nor to anything said or done at the meeting. Still, there is a difference in degree, if not in kind, between events like the department's meeting and events like Oliver. So perhaps we should go along with Lewis here; perhaps we should deny that there are any events which are essentially non-occurrences of other events. But this restriction will turn out to prove fatal to Lewis's theory of causation.

Failures As Causes

Tom didn't water the plant the week that Sally was away. That's a fact. The plant died; that's also a fact. Tom was supposed to water the plant; Sally told him it would die for sure if he didn't. But Tom forgot and the plant died.

Tom's failure to water the plant caused its death. This is an uncontroversial causal claim, one that Lewis would like to accommodate. The most natural way to do this would be to extend his counterfactual analysis to facts. This fact-event counterfactual is true:

(12F) If Tom had watered the plant, its death would not have occurred.

Lewis can't extend his analysis to fact causation because, as we've seen, there are true backtracking fact counterfactuals. But there is a way of counting this as a case of event causation. People who think that causation is a relation between events and *only* between events but who nevertheless want to count some failures as causes recognize a class of events they call 'events of omission'. If Lewis wants to count Tom's not watering the plant as the cause of its death, he could say that Tom's failure to water the plant is his omission to do so. And he could say that Tom's omission to water the plant caused its death because the appropriate *event* counterfactual is true:

(12E) If Tom's omission to water the plant had not occurred, its death would not have occurred.

But is (12E) true? That depends on what the event of Tom's omission is.

If we think that omissions are actions and that actions are bits of behavior with the appropriate rational causes, we may be tempted to say that Tom's omission is whatever Tom *does* when he should have been watering the plant.

But this won't do. Suppose that Tom plays a marathon game of poker which lasts for the whole week that Sally is away. Then on this view Tom's omission to water the plant *is* his playing poker. But *this* event (the poker game) may fail to occur without it being true that Tom waters the plant. And it may be that the closest worlds where Tom's poker game fails to occur are worlds at which he still forgets to water the plant and where it still dies. So if Tom's plant-watering omission *is* his playing poker, (12E) should be read as (12E'), and (12E') is false:

(12E') If Tom's playing poker had not occurred, the plant's death would not have occurred.

The point is this: When we evaluate counterfactuals about what would have happened had an event not occurred, we must be careful to do so by considering the essential properties of the event, not by considering its accidental properties. Granted, at the actual world Tom fails to water the plant *by* spending the week playing poker at Fred's house, and granted that someone might say that Tom's omission and his poker game are one and the same event under different descriptions. But being a playing of poker is an accidental, not essential, property of Tom's failure to water the plant. There are other nearby worlds where Tom fails to water the plant by spending the entire week writing a paper on causation, working so hard that he forgets to feed the cat, let alone water the plant. (Tom's an obsessive kind of guy.) So *if* Tom's omission to water the plant is an event which caused the death of the plant, it is *not* an event which is essentially his playing poker.

If we want to get (12E) to come out true, then we must count Tom's omission to water the plant as an event which is *essentially* an omission. That is, we must count Tom's omission as an essentially negative event which occurs regardless of *how* he fails to water the plant and which fails to occur just in

case he *does* water the plant. To avoid confusion, let's call Tom's omission, thus understood, 'Tom's failure to water the plant'. Then it *is* true that:

(12E") If Tom's failure to water the plant had not occurred, its death would not have occurred.

And Lewis can say that Tom's failure to water the plant caused its death.

But if Lewis says this, he is in trouble. If Lewis counts Tom's failure as an event, he has no grounds for excluding events like Oliver. Lewis's rationale for excluding events like Oliver was that such events are disjunctive in 'overly varied ways'; Oliver is an event which fails to occur if and only if Kennedy is inaugurated for a second term as President. But Tom's failure to water the plant is also an event which is disjunctive in highly varied ways and which fails to occur just in case Tom waters the plant.

Lewis faces a dilemma: If he counts Tom's failure as an essentially negative, highly disjunctive event, then he must count Oliver as an event. But if he counts Oliver as an event, the Event Asymmetry thesis is false, and Lewis's account fails to distinguish causes from effects. So Lewis *cannot* count Tom's failure as an event. But if he denies that Tom's failure is an event, then his account of causation cannot handle failures as causes.¹⁶

The Difference Between Causes and Effects

Lewis needs to ban events like Bob and Oliver because his account of the causal relation makes it too easy for two events to be causally related. His ban on events with overly detailed essences (like Bob) and on events with overly disjunctive essences (like Oliver) preserves the truth of the Event Asymmetry thesis, but the cost is that he lacks enough causal *relata* to handle all cases of causation.

The advantage of my more complex account of the causal *relation* is that I need fewer restrictions on what counts as causal *relata*. In particular, I don't need to ban events like Bob and Oliver. This gives my account a flexibility Lewis lacks -- I can appeal to highly fragile events to draw needed distinctions in overdetermination cases. And I can count failures as causes.

Recall that on my view an event C caused an event E just in case it's true that:

- (4) If C had not occurred, then (E would not have occurred *and* if C had occurred, then E would have occurred).

Now consider Tom again, and suppose that his failure to water the plant is an essentially negative event which fails to occur just in case he waters the plant. Then on my account it's true that his failure caused the death of the plant if and only if it's true that:

- (13) If Tom's failure to water the plant had not occurred, then (the plant's death would not have occurred *and* if Tom's failure to water the plant had occurred, its death would have occurred).

Or, more naturally, if this equivalent fact-event counterfactual is true:

- (14) If Tom had watered the plant, then (its death would not have occurred *and* if he had not watered it, its death would have occurred)

That is, on my account it's true that Tom's failure to water caused the death of the plant just in case the closest worlds at which he waters the plant are worlds at which it's true not only that its death doesn't occur, but also true that its death *would* have occurred had he *not* watered it. To put it another way, on my account Tom's failure to water the plant counts as a cause of the plant's death just in case the plant's life or death (that week) depends on whether or not Tom waters it.

Now consider the problem of distinguishing causes from their effects. Lewis had to say that Oliver is not an event because if Oliver is an event, this backtracking event-event counterfactual is true:

- (11) If Oliver had not occurred, then Kennedy's assassination would not have occurred.

and the truth of (11) commits Lewis to the claim that Oliver is a backwards cause of Kennedy's assassination.

But I don't need to ban events like Oliver. For on my account, the truth of a backtracking event-event counterfactual like (11) does not suffice for the truth of a backwards causal claim. On my view, it's true that Oliver

caused Kennedy's earlier assassination only if it's true that (11) and *also* true that:

- (15) If neither Oliver nor Kennedy's assassination had occurred, then (if Oliver had occurred, Kennedy's assassination would have occurred.)

That is, only if it's true that:

- (16) If Kennedy's assassination had not occurred and he had been inaugurated for a second term, then (if his second inauguration had *not* occurred, his assassination would have occurred).

And (16) is false. For consider the closest worlds where Kennedy's assassination didn't happen and where he was inaugurated for a second term as President. Now ask this question: Is it true, at *these* worlds, that if Kennedy's second inauguration had *not* occurred - then his assassination would have occurred? Obviously not. If Kennedy were alive and well and being inaugurated in 1965, then it's false that if he were not being inaugurated, he would have been killed over a year earlier.

What about events like Bob? Recall that Lewis has to deny that Bob is an event because if Bob is an event, then this backtracking event-event counterfactual is true:

- (9) If the driver's death had not occurred, then Bob would not have occurred.

and the truth of (9) commits Lewis to the claim that the driver's death is a backwards cause of the big and fragile event which is Bob.

But I can count Bob as an event and admit that (9) is a true backtracking event-event counterfactual without saying that the driver's death caused Bob. For on my view, the driver's death caused Bob only if it's true that (9) and *also* true that:

- (16) If the driver's death had not occurred, then (Bob would not have occurred and if his death had occurred, then Bob would have occurred).

That is, only if it's true that:

- (17) If the driver's death had not occurred, then (something or other would have happened differently from the time of the accident until just before his death *and* if his death had occurred, then everything would have happened exactly as it actually happened).

And (17) is not true. Consider the world at which the driver swerves to the left rather than to the right, and is injured but survives. At this world, Bob fails to occur. Now ask: Is it true, at this world, that if the driver's death had occurred, he would have swung to the right *and* everything else would have happened exactly as it actually happened? Surely not. But then it's false, at this world, that if the driver's death had occurred, Bob would also have occurred. Bob is a very fragile event; this means that it's very easy for Bob not to occur, and correspondingly difficult for Bob to occur.

I've defended my theory of causation by showing that it succeeds where Lewis fails. My account can handle failures as causes, and it distinguishes causes from effects without relying on a gerrymandered theory of events. There is obviously a great deal more to be said; I need to explain why the counterfactual relation expressed by (4) is contingently temporally asymmetric, and I need to provide an account of causal relata. I cannot do this here. Instead, I will conclude by claiming another virtue for my account. Throughout this paper, I used the language of event causation. I did this mostly for strategic purposes, to better contrast my account with Lewis's. But my account, unlike Lewis's, does not require a distinction between events and facts. If you think that causal relata are facts or states of affairs, then my account can be adapted to this. I would also argue, though not here, that if I am right about what the causal relation is, there is less to the fact/event distinction than we might have thought.¹⁷

University of Southern California

FOOTNOTES

- 1 Of course, as Davidson has taught us, it need not be a law that we know. The general fact need not be a fact about logs and fires; all that's required is that the log-adding and the fire's burning instantiate some properties F and G such that F events are nomologically connected to G events. ('Causal Relations' *Journal of Philosophy* 64 (1967) pp. 691-703.)
- 2 See, for instance, Hilary Putnam, 'It Ain't Necessarily So', *Journal of Philosophy* 59 (1962) pp.658-671, and David Lewis, 'The Paradoxes of Time Travel', *American Philosophical Quarterly* 12 (1976) pp.145-152.
- 3 D. Lewis, 'Causation', *Journal of Philosophy* 70 (1973) pp.556-567. Also in D. Lewis, *Philosophical Papers, Volume II* (Oxford: Oxford University Press, 1986) pp.159-213.
- 4 The Regularity theory and Lewis have opposite problems in cases where there are two events A and B such that A in fact caused E, but B would have caused E had A not pre-empted B as a cause of E. In some of these cases, (cases where A and B are both spatiotemporally contiguous with E and where there are no intermediate events on which E counterfactually depends), the Regularity theory falsely implies that B caused E and Lewis's theory falsely implies that A did not cause E.
- 5 Despite his claims, in 'Counterfactual Dependence and Time's Arrow', *Nous* 13 (1979) pp.455-476, Lewis does not offer us such an account either. In that paper, Lewis defends a similarity ranking for counterfactuals which, together with a contingent feature of the world (the overdetermination of the past by the future), results in a contingent temporal asymmetry of counterfactuals. The closest worlds, according to Lewis, are those at which most of the past is the same as our past history, but at which the future is often quite different. But this does not solve the problem of distinguishing causes from effects, because Lewis allows that the closest worlds may differ from ours with respect to the immediate past.
- 6 This is rough. Lewis in fact defines causation in two steps; causal dependence is counterfactual dependence as defined by (1) and (2); A caused B iff either B causally depends on A or B is linked to A by a chain of causal dependences. I will ignore this complication, since none of my criticisms of Lewis turns on this distinction. ('Causation', *ibid.*)
- 7 D. Lewis, *Counterfactuals* (Cambridge: Harvard University Press, 1973).

-
- 8 Although Lewis occasionally says things which suggest otherwise, he holds that we standardly evaluate counterfactuals by considering worlds at which past history diverges from our history at some time before the time of the antecedent. This entails that there are standardly true backtracking counterfactuals. ('Counterfactual Dependence and Time's Arrow, *ibid.*)
- 9 Interestingly, Lewis never actually says this. What he in fact says is that 'there may be no true counterfactuals that say in any detail how the immediate past would be if the present were different'. ('Counterfactual Dependence and Time's Arrow', *Nous* 13 (1979) p.463.) Jonathan Bennett has persuaded me that what Lewis meant to say is that for any two events e_1 and e_2 , there is no standardly true counterfactual which implies that if e_2 had not occurred, then e_1 would not have occurred. ('Counterfactuals and Temporal Direction', *Philosophical Review* 93 (1984) pp.57-91.)
- 10 'Events', *Philosophical Papers, Volume II* (Oxford: Oxford University Press, 1986) pp. 241-270.
- 11 'Events', *ibid.*, pp.255-256. See also pp.258-260 for Lewis's account of why the event of someone's writing 'rr' and the event of his writing 'Larry' are not wholly distinct events.
- 12 'Events', *ibid.*, p.250.
- 13 'Causation', *Philosophical Papers, Volume II, ibid.*, p.204.
- 14 'Events', *ibid.*, p.266.
- 15 'Events', *ibid.*, p.250.
- 16 Lewis realizes he is in trouble with omissions as causes. His brief discussion in Postscript D to 'Causation' (*Philosophical Papers, Volume II, ibid.*, 189-193) ends with his admission that omissions as causes leave him with 'unfinished business'.
- 17 This paper has been a long time in the writing. It began as a postscript to a paper on the problem of free will and determinism ('Freedom, Causation, and Counterfactuals' *Philosophical Studies* 64 (1991) pp.161-184); earlier versions were presented to philosophers at the University of California at Riverside, Trinity University, Dalhousie University, the Atlantic Philosophy Association meetings (1993), the Northwest Philosophy conference (1993), and the Pacific Division American Philosophical Association meetings (1994). I am grateful to audience members for helpful comments, and especially to Bob Bright and Terrance Tomkow.