Counterfactuals and Preemptive Causation

By

Jonardon Ganeri, Paul Noordhof, Murali Ramachandran

David Lewis modified his original theory of causation in response to the problem of 'late preemption' (see 1973b; 1986b: 193-212). However, as we will see, there is a crucial difference between genuine and preempted causes that Lewis must appeal to if his solution is to work. We argue that once this difference is recognized, an altogether better solution to the preemption problem presents itself.

1. *The Preemption Problem and Lewis's Solution*

According to Lewis's original theory, for any actual, distinct events *c* and *e*, *e depends* (counterfactually) on *c* just in case it is true that if *c* had not occurred, then *e* would not have occurred; and *c causes e* if and only if

(L) there is a series of (actual) events $X_1, ..., X_n$ such that: X_1 depends on *c*, X_2 on X_1, ..., and *e* depends on X_n.

When (L) is met, let us say *c* is an *ancestor* of *e* or, alternatively, that *e* is a *descendant* of *c*. It is the thesis that *c* causes *e only if c* is an ancestor of *e* that comes under threat from the possibility of late preemptive causation.[1] Here's one case. Suppose two marksmen, Patel and Singh, are targeting a balloon at a fairground stall; they pull their

[1] Lewis accommodates simpler cases of preemption by appeal to the general asymmetry of counterfactual-dependence (see 1979: 33-34; 1973b: 170-72). But he recognizes that this strategy does not work for cases of late preemption, such as the one we are about to consider.

respective triggers virtually simultaneously; Patel's pellet reaches the balloon a split second before Singh's and bursts it; but, had Patel missed, or simply been a fraction slower, Singh's pellet would have burst the balloon instead. Now, in virtue of what does Patel's action (*P*), *but not* Singh's action (*S*), cause the bursting of the balloon (*B*)? The fact is, there is no descendant of *P* upon which *B* counterfactually depends. *B* does not even depend on the impact between Patel's pellet and the balloon (*K*), for example; for if *K* had not occurred, then *Singh's* pellet, by hypothesis, would have made impact and caused *B* instead.

So, what gives?

Lewis. He embellishes his original account with the notion of *quasi-dependence* (1986b: 205-207) and redefines causation as follows: for any actual events *c* and *e*, *c* causes *e* just in case

(L*) there is a series of (actual) events X_1, ..., X_n such that: X_1 depends or quasi-depends on *c*, X_2 depends or quasi-depends on X_1, ..., and *e* depends or quasi-depends on X_n.

Thus, in particular, *c* causes *e* if *e* merely quasi-depends on *c*. Here's a rough idea of what quasi-dependence amounts to and how it's meant to resolve the problem. In our example, the preempting process stemming from *P* is *independent* of the process stemming from *S*, in that it evidently is possible (in our world or in other worlds with the same laws) for a process exactly similar to the former to occur, along with an event exactly like *B*, without the preempted process occurring at all. Were such a possibility to obtain, the *B*-similar event (*B**) presumably *would* depend on the *P*-similar event (*P**); i.e. it would be true that if *P** had not occurred, then *B** would not have occurred. *B* *quasi-depends* on *P* if the vast majority of processes that are exactly similar to the process between *P* and *B*, as measured by variety of surroundings, likewise constitute *bona fide* causal chains (1986b: 206). In that case, given (L*), we get the desired result

that *P* causes *B*.

But, what seems to have gone unnoticed is that, as it stands, this strategy does not rule out *S* as a cause of *B*! For, it is also possible (in our world or in other worlds with the same laws) for a process exactly like the *S*-process to occur, along with an event just like *B*, without the *P*-process occurring at all. Indeed, this possibility is acknowledged in the very characterisation of a preemption situation: if *c*, the preempting cause, had not occurred, then *d*, the preempted cause, would still have occurred and would have caused *e*. Were this possibility to obtain, then, the *B*-similar event *would* counterfactually depend on the *S*-similar event. So, by parity of reasoning, we are led to the conclusion that *B* quasi-depends on *S*; and thence, by (L*), to the *undesired* conclusion that *S* causes *B* as well.

So far as we can see, Lewis can avoid this conclusion only by denying that the processes stretching from *S*-like events to *B*-like events which exhibit the proper pattern of counterfactual dependence are *exactly like*, or have the same 'intrinsic character' as, the actual *S*-process. This denial is warranted in so far as these counterfactual processes include an event that does not have a counterpart in the actual *S*-process: namely, the impact between Singh's pellet and the balloon. There is no corresponding mis-match between the actual *P*-process and its cousins. Thus, this difference -- and, in our opinion, *only* this difference -- between the *P*- and *S*-processes allows Lewis to maintain that *P*, but not *S*, causes *B*.

However, the same difference also suggests an alternative resolution of the whole problem, including an alternative theory of causation.

2. *The Alternative Solution*

Consider our balloon example again. Instead of appealing to the (supposed) fact that processes *similar* to the *P*-*B*-process (the process leading from *P* to *B*) normally, or in the vast majority of cases (or whatever), do exhibit the proper pattern of dependence, we

propose to play on the basic fact that *P* itself is a *potential* ancestor of *B*: if *P* had occurred *without S* occurring, then *B might have* occurred and been a descendant of *P*. We say *B* '*might have*' occurred rather than '*would have*' so as to allow 'accidental' causation. Suppose Patel is *not* a marksman and that he bursts the balloon by fluke (maybe he sneezes just as he is pulling the trigger). We still want to say that his pulling of the trigger (*P*) causes the bursting of the balloon (*B*), but not that *P would have* resulted in the occurrence of *B* if *S* had not occurred; for, since his success is a fluke, there will be near-enough worlds where *P* occurs without *B*. It does seem legitimate, however, to say that *B might have* occurred if *P* had occurred without *S*.[2]

Of course, *S* is also a potential ancestor of *B* in this sense: i.e. it is also true that if *S* had occurred without *P* occurring, then *B* might have occurred and been a descendant of *S*. But here we can resort to the crucial difference between the *P*- and *S*-processes mentioned in the last section. We can avoid the conclusion that *S* causes *B* by appealing to the fact that the *S*-*B*-process is 'incomplete', in that: *were S* to be an ancestor of *B*, there would be a *non-actual* event (i.e. an event that does not occur in the actual world) in *that S*-*B*-process. The required impact between Singh's pellet and the balloon is one such non-actual event. The *P*-*B*-process, on the other hand, is 'complete': *were P* to be an ancestor of *B*, the *P*-*B*-process *might* be (*could have* been) comprised *solely of actual events*.

These suggestions give rise to the following analysis of causation. Let us say an event *k mediates* between events *c* and *e* at world *w* if and only if it is true at *w* that *c* is an ancestor of *k* and that *k* is an ancestor of *e*. Intuitively, if *k* mediates between *c* and *e*, *k* is an event in the causal process leading from *c* to *e*. Next, let us stipulate that when we use the phrase 'actual event' in characterising a counterfactual situation, we mean an

[2] Following Lewis (1973a), we take conditionals of the form, 'If *A* were the case, *B might* be the case' (symbolically: $A \diamond\!\!\rightarrow B$) to be equivalent to 'It is not true that: if *A* were the case, then *B would not* be the case' (symbolically: $\neg(A \Box\!\!\rightarrow \neg B)$).

event that occurs in the *actual* world, *this* world *we* inhabit, *not* the world that is 'actual' as far as that counterfactual situation is concerned -- we have already adopted this use of the phrase in the previous paragraph. Here, then, is the analysis we wish to promote. For any actual, distinct events *c* and *e*, *c* causes *e* if and only if

> **PCA** there is a (possibly empty) set of actual events, Σ, such that if *c* were to occur *without* any of the events in Σ, then *e might* occur as a descendant of *c* with only actual events mediating between *c* and *e*.

This analysis counts a course of events as a *causal process* if (very roughly) these events *might have* formed a chain of counterfactually-dependent events but for some actual events and done so without the 'aid' (more specifically, the mediation) of any non-actual events. Speaking loosely, we may say causes are 'potentially complete ancestors' of their effects. Since no suitable name suggests itself, we shall simply refer to our theory as the *PCA*-analysis of causation.

By the PCA-analysis, if *c* is an ancestor of *e*, it is a cause of *e* (as is the case with Lewis's theory). Just take Σ to be the empty set. It is trivially true that PCA is satisfied. For, the actual world itself is the nearest world to actuality where *c* occurs, and, by hypothesis, it is a world where *e* is a descendant of *c*; and the mediation condition, as we might call it, is also met, since every event that occurs in the actual world is obviously an actual event! Hence, *c* comes out as causing *e*, as required.

And it is easy to confirm that we get the right verdicts in our balloon example. Patel's pulling of the trigger (*P*) comes out as a cause of the bursting of the balloon (*B*); just take Σ to be {*S*}; if *P* were to occur without *S*, then it is surely possible for the actual *P*-*B*-process to occur in such circumstances and form a 'complete' causal chain. As for Singh's pulling of the trigger (*S*), it is true that if *S* were to occur without *P*, then *B* might occur as its descendant; but *S* is ruled out as a cause of *B* by the mediation requirement: as we have noted, the potential impacts between Singh's pellet and the

balloon that would have to occur for *S* to be an ancestor of *B* do not occur in the actual world.

So, we have an alternative analysis of causation that also resolves the preemption problem. The question arises, why should one prefer the new account to Lewis's?

3. *Comparisons with Lewis's Solution*

We offer three reasons why Lewisians themselves should favour our account over the quasi-dependence account.

(1) If, as Lewis maintains (1986b: 205-206), a course of events constitutes a causal process *only* in virtue of the 'intrinsic' features of that process, a theory that omits to identify these specific features cannot be regarded as capturing the essential character of causation. The quasi-dependence account is surely such a theory. It picks out causal processes *indirectly* in that it deems a process causal by way of its intrinsic similarity to other processes that *do* exhibit the proper pattern of counterfactual dependence. Our account, on the other hand, deems a process causal by way of *its* counterfactual properties alone. So it can at least lay claim to specifying the very features *in virtue of which* one event is a cause of another.

It may be objected that the contrast between Lewis's account and ours in this regard is exaggerated. Since the truth of counterfactual conditionals is determined by how things stand in other possible worlds, the objection runs, one event is a cause of another in virtue of how things stand with *counterparts* of those events in those worlds. In that case, however, the PCA-analysis too must be regarded as appealing (albeit indirectly) to events that are merely *similar* to (rather than *identical* with) the putative cause and effect. So, the objection concludes, there is no substantive difference between the proposed solution to the preemption problem and that offered by the quasi-

dependence account.[3] Now, it is true that Lewis (e.g. 1986d: ch. 4) appeals to *counterparts* of objects in evaluating *de re* modal statements; this is because, according to him, no object exists in more than one possible world. But, the present objection mistakenly takes Lewis to be committed to a similar view of *events*. He isn't. Events, unlike individuals, can exist in more than one world in his metaphysics; witness:

> To any event there corresponds a property of regions: the property that belongs to all and only those spatiotemporal regions, *of this or any other possible world*, in which that event occurs [...] The property that corresponds to an event, then, is the class of all regions—at most one per world—*where that event occurs*. (1986c: 244, my emphases)

So, as far as Lewis is concerned, the PCA-analysis may well be seen as identifying the counterfactual-relations that must hold between two events *themselves* for one to be a cause of the other. The quasi-dependence analysis cannot make the same claim.

(2) A second advantage of our account is that it does not assume, as Lewis's account does, that: *if two processes are intrinsically similar, then either both are causal or neither are*. Quite apart from the fact that the notion of '*intrinsic* similarity' is never clearly defined by Lewis, the problem with the above assumption is that it rules out the possibility of brute singular causation, of *one-off* causation between particular events. This should concern Lewis since he originally strove to accommodate this possibility; for example, he decides against repaired regularity analyses on the grounds that they 'would gratuitously rule out inexplicable causal dependence, which seems bad' (1973b: 169, fn. 11). Since the PCA-account focuses exclusively on the counterfactual properties of a process itself in determining whether it is causal, the account does not rule out

[3] We are grateful to an anonymous referee for this point.

primitively singular causation a priori.

(3) Finally, since the PCA-analysis calls on nothing more than the counterfactual-dependence relation, it can simply borrow, as Lewis's original account (1973b) does, the semantics and logical framework behind his theory of counterfactuals (1973a). So, it is an easy matter to provide a formal semantics for our account (we pave the way in the appendix). The formalization of the quasi-dependence account, however, calls for the assimilation of notions like 'process' and 'intrinsic similarity', which, on the face of it, are irreducible; at any rate, a suitable model seems some way off.

We submit, then, that our account is preferable to Lewis's quasi-dependence account because it handles preemptive causation just as well but better captures Lewis's initial intentions and lends itself to a simpler formal semantics.[4]

Jonardon Ganeri; Paul Noordhof
University of Nottingham
Nottingham NG7 2RD
Jonardon.Ganeri@nottingham.ac.uk
Paul.Noordhof@nottingham.ac.uk

Murali Ramachandran
School of English & American Studies
University of Sussex
Falmer, East Sussex, BN1 9QH
muralir@sussex.ac.uk

[4] **This footnote does not appear in the original article.** The analysis presented here is modified slightly in Ganeri, Noordhof, and Ramachandran, 'For a (revised) PCA-analysis', *Analysis* 58 (1998) pp. 45-47. This is in response to objections put forward in A. Byrne and N. Hall, 'Against the PCA-analysis', *Analysis* 58 (1998) pp. 38-44.

References

Lewis, D. 1973a. *Counterfactuals*. Oxford: Basil Blackwell.

Lewis, D. 1973b. Causation. *Journal of Philosophy* 70: 556-67. Reprinted in his 1986a: 159-72, to which references apply.

Lewis, D. 1979. Counterfactual Dependence and Time's Arrow. *Noûs* 13: 455-76. Reprinted in his 1986a: 32-52.

Lewis, D. 1986a. *Philosophical Papers*, Volume 2. Oxford: OUP.

Lewis, D. 1986b. Postscripts to 'Causation'. In his 1986a: 172-213.

Lewis, D. 1986c. Events. In his 1986a: 241-69.

Lewis, D. 1986d. *On the Plurality of Worlds*. Oxford: Basil Blackwell.

Appendix: A formal characterisation of the PCA-analysis

In the following definitions, *A* and *B* are arbitrary propositions, *c*, *d* and *e* are arbitrary possible events, and the quantifiers range over possible events. We assume the logical framework of Lewis's (1973a) theory of counterfactuals—so the reader should look there, for example, for the treatment of 'closeness' of, or similarity between, possible worlds.

Df1: w^* $=_{df}$ The actual world.

Df2: $\vdash\!\!\!-_w A$ $=_{df}$ *A* holds in world *w*

Df3: $\vdash\!\!\!-_w @(A)$ $=_{df}$ $\vdash\!\!\!-_{w^*} A$

Df4: $\vdash\!\!\!-_w O(e)$ $=_{df}$ *e* occurs in world *w*

Df5: $\vdash\!\!\!-_w A \,[]\!\!\rightarrow B$ $=_{df}$ Some *A*-world where *B* holds is closer to world *w* than any *A*-world where *B* does not hold. (Informally: it is true in *w* that if *A* were the case, then *B* would be the case.)

Df6: $\vdash\!\!\!-_w A \Diamond\!\!\rightarrow B$ $=_{df}$ $\vdash\!\!\!-_w \neg(A \,[]\!\!\rightarrow \neg B)$
(Informally: it is true in *w* that if *A* were the case, *B might* be the case.)

Df7: $\vdash\!\!\!-_w c \supset\!\!\rightarrow e$ $=_{df}$ *c* is an *ancestor* of *e* in world *w*.

$=_{df}$ $\vdash\!\!\!-_w \exists x_1 \ldots \exists x_n(O(x_1) \,\&\, \ldots \,\&\, O(x_n) \,\&\, (\neg O(c) \,[]\!\!\rightarrow \neg O(x_1)) \,\&\, \ldots \,\&\, (\neg O(x_n) \,[]\!\!\rightarrow \neg O(e)))$

Df8: $\vdash\!\!\!-_w \mathbf{M}dce$ $=_{df}$ *d mediates* between *c* and *e* in world *w*.

$=_{df}$ $\vdash\!\!\!-_w (c \supset\!\!\rightarrow d) \,\&\, (d \supset\!\!\rightarrow e)$

THE PCA-ANALYSIS OF CAUSATION

For any actual events *c* and *e*, *c* causes *e* iff

PCA there is a (possibly empty) set of actual events, Σ, such that:

$\vdash\!\!\!-_{w^*} (O(c) \,\&\, \forall x((x \in \Sigma) \rightarrow \neg O(x))) \Diamond\!\!\rightarrow (O(e) \,\&\, (c \supset\!\!\rightarrow e) \,\&\, \forall x(\mathbf{M}xce \rightarrow @O(x)))$