Against the PCA-analysis

Alex Byrne & Ned Hall

Jonardon Ganeri, Paul Noordhof, and Murali Ramachandran (1996) have proposed a new counterfactual analysis of causation. We argue that this – the *PCA-analysis* – is incorrect. In section 1, we explain David Lewis's first counterfactual analysis of causation, and a problem that led him to propose a second. In section 2 we explain the PCA-analysis, advertised as an improvement on Lewis's later account. We then give counter-examples to the necessity (section 3) and sufficiency (section 4) of the PCA-analysis.

1. Lewis's two analyses and the problem of late preemption

Following Lewis, say that event *e counterfactually depends* on event *c* iff, if *c* had not occurred, *e* would not have occurred. In his 1973, Lewis proposed that *c* causes *e* iff *e* is related to *c* by the ancestral of the relation of counterfactual dependence. (Following Ganeri et al., say that *e* is a *descendant* of *c*, and *c* is an *ancestor* of *e*, iff Lewis's analysans holds.) However, this analysis does not give a *necessary* condition for *c* to cause *e*, as figure 1 shows:

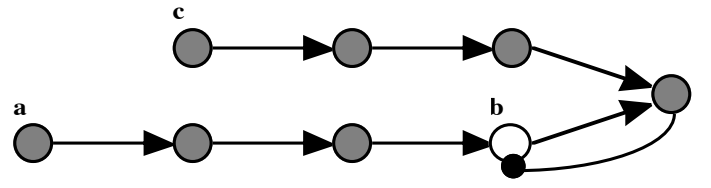

Figure 1

Here the circles represent 'neurons'; an arrow connecting two neurons shows that the firing of the first causes the firing of the second; replacing the arrowhead by a blob shows that the firing of the first inhibits the firing of the second, even if the latter is stimulated by other neurons; and filled circles represent that the corresponding neuron fires (such helpful diagrams are introduced in Lewis 1986b).

Figure 1 depicts *c* (i.e. the event of neuron **c**'s firing) causing *e*, which then 'preempts' the causal process that begins with *a*, by inhibiting neuron **b**. But *e* is not a descendant of *c*: there is no series of events $X_1, \ldots, X_n$ such that X_1 counterfactually depends on *c*, X_2 on X_1, …, and *e* on X_n. Thus Lewis's analysis fails to give a necessary condition for causation.

A *process* is a course of events (Lewis 1986b: 205). *e quasi-depends* on

Analysis 58.1, January 1998, pp. 38–44.

c iff some process that has *c* as its first member and *e* as its last member is 'in its intrinsic character … just like processes in other regions (of the same world, or other worlds with the same laws) situated in various surroundings' and 'the great majority – as measured by variety of the surroundings' of these processes are such that their last members counterfactually depend on their first members (206).

Armed with these definitions, Lewis amended his analysis of causation as follows: *e causally depends* on *c* iff it either counterfactually depends on *c*, or quasi-depends on *c*; *c* causes *e* iff *e* is related to *c* by the ancestral of the relation of causal dependence (206–7). And this revised analysis does appear to overcome the problem posed by figure 1. Intuitively, *e* quasi-depends on *c*: a process intrinsically like the *c-e* process, in surroundings that did not involve extraneous preempted processes leading to *e*, would be one in which *e* counterfactually depended on *c*; furthermore, keeping the laws fixed, 'the great majority' of possible duplicate *c-e* processes will have this feature.

2. *The PCA-analysis*

Although Ganeri et al. take Lewis's revised analysis to be extensionally correct, they complain that, inter alia, it 'calls for the assimilation of notions like "process" and "intrinsic similarity", which, on the face of it, are irreducible' (1996: 224).[1] Their analysis is designed to avoid such extra theoretical machinery. Following Ganeri et al., 'say that 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*' (221). Then the PCA-analysis is this:

> *c* causes *e* iff '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*' (222).

[1] In the case of 'process', their complaint is not warranted: as we mentioned a few paragraphs back, Lewis defines a process to be a course of events. However, there is a complaint to be made, because the official account of a process sits ill with other claims Lewis makes. For example, he says that there is some 'slight presumption' in favour of the view that the causal structure of a process is fixed by its intrinsic character, plus the laws (1986b: 205). But this, far from being intuitive, is in fact false if we take a process to be a course of events. (Suppose a particle emission *a* occurs at *t* which causes a particle collision *b* at $t+d$. Keeping the laws fixed, a duplicate of the course of events consisting of *a* and *b* might well not have the same causal structure – imagine the emitted particle gets knocked off course but a duplicate of *b* is caused to occur, by other means, *d* time units later than the duplicate of *a*.) For more discussion, see Hall 1997.

The intuitive idea is that a series of events forms a causal chain iff, but for the occurrence of some extraneous events, it might have been that each event in the series was counterfactually dependent on its predecessor. 'Speaking loosely, we may say that causes are "potentially complete ancestors" of their effects' (222).

If, in the figure 1 situation, *c* had occurred without *a*, then *e* would (and so *might*) have occurred as a descendant of *c* with only actual events (i.e. those occurring according to figure 1) mediating between *c* and *e*. Applying the PCA-analysis to figure 1, and taking Σ to be {*a*}, we get the correct result that *c* causes *e*.

And although if *a* had occurred without *c*, *e* would have occurred as a descendant of *a*, it would *not* have occurred as a descendant of *a* with only actual events mediating between *a* and *e*: in such a situation, neuron **b** would have fired. Thanks to the 'only actual events mediating' clause, the PCA-analysis does not wrongly classify *a* as a cause of *e*.

So far, so good.

3. *Failure of necessity*

The PCA-analysis does not give a necessary condition for *c* to cause *e*, as figure 2 shows:

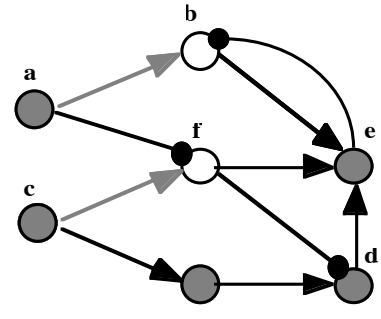

Figure 2

Here, the shaded arrows indicate action at a temporal distance: there are no events that carry the stimulatory signal from **a** to **b** or from **c** to **f**; rather, the laws simply state that if **a** fires, then a certain time later **b** will fire (unless it has been inhibited). Likewise for **c** and **f**.[2]

The timing of events represented by this diagram is important. Here's how it goes: **a** and **c** fire simultaneously, at time 0; **d** fires at time 2; **e** fires at time 3. Had **a** not fired, **f** would have fired at time 1, causing **e** to fire at

[2] Lewis (1986b: 202–3) suggests that cases involving action at a temporal distance are spoils to the victor. We disagree: many of them (including the one discussed here) seem to us to be perfectly clear examples of causation.

time 2 and preventing **d** from firing at time 2. The time delay between *a* and *b* is long: had **b** not been inhibited, it would have fired at time 4 as a result of *a*, in turn causing **e** to fire at time 5.

e is not a descendant of *c*: pick any event in the *c-d-e* chain, and had that event not occurred, *e* would still have occurred (albeit at time 5, not at time 3[3]). But *c* causes *e*, so if the PCA-analysis is right, there must be some set of events Σ such that, had *c* occurred without any events in Σ occurring, *e* might have occurred as a descendant of *c*, with only actual events mediating between *c* and *e*.

Clearly, Σ must include *a*. (Had there been a stimulatory signal between **a** and **b**, Σ could have included one of its constituent events instead.) Furthermore, surveying the events of figure 2, we see that Σ must *not* include any of these events *besides a*. For the only other events are those constituting the *c-d-e* chain; had *c* occurred without one of *them*, then *e* would not have occurred as a descendant of *c* at all – or at any rate, even if it *did*, non-actual intermediates would have to be involved. (Again, had there been a stimulatory signal between **c** and **f**, Σ could have included one of its constituent events as well.)

So the only remaining candidate for Σ is {*a*}. But if *c* had occurred without *a*, the sequence of events would have been as depicted in figure 3:

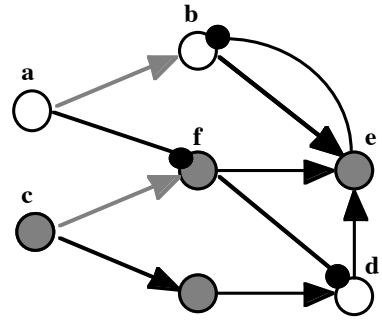

Figure 3

To be sure, *e* occurs as a descendant of *c* – but with non-actual intermediates aplenty. According to the PCA-analysis, then, *c* does not cause *e*, and thus the analysis does not give a necessary condition.[4] (Notice that this example does not pose any special trouble for Lewis's quasi-dependence account.)

[3] Would such a firing of **e** at time 5 have *been e*? To say no would be to adopt an extremely 'fragile' conception of events. Such a conception would solve the problem of late preemption straightaway, but as Lewis (1986b) argues, the cure is worse than the disease.

[4] A fix, in the spirit of the PCA-analysis, is this. Weaken the condition for *c* to cause *e* by allowing the counterfactual situation in which *c* occurs to contain, not only the

4. *Failure of sufficiency*

Without the 'only actual events mediating' clause, the PCA-analysis would certainly fail to supply a sufficient condition (e.g., *a* would be misclassified as a cause of *e*, in figure 1). Suppose we have a case where *c* does not cause *e*, but would have, but for the occurrence of some event *a*. More specifically, had *c* occurred but *a* not, then *e* would have occurred, and would have been a descendant of *c*. Enter the 'only actual events mediating' clause: in such a counterfactual situation – according to the PCA-analysis – there is some non-actual event *d* that mediates between *c* and *e*.

The problem of late preemption shows us how easy it is to construct examples in which some event *d causally* mediates between *c* and *e* (i.e. is an effect of *c* and a cause of *e*), but does not *mediate* between *c* and *e*. And, as should come as no surprise, such cases spell trouble. For consider figure 4:

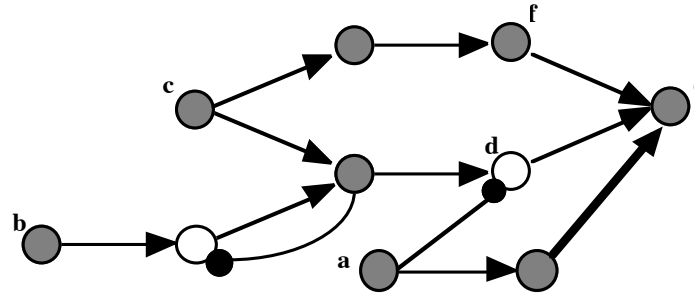

Figure 4

Here **e** is a neuron that needs a specially strong stimulation in order to fire, and it gets that from **a**. But the stimulation **e** receives from **f** is too feeble on its own to cause **e** to fire. However, stimulation from **f** and **d** (which does not actually fire), although individually insufficient, would be jointly sufficient for **e** to fire. Clearly, *c* is not a cause of *e*. However, had *c* occurred but *a* not, the sequence of events would have been as depicted in figure 5:

---

absence of certain actual events, but also the presence of certain actual non-events, or 'omissions'. (Then, in the figure 2 case, we can consider a counterfactual situation in which *a* does not occur and **f** does not fire.) More exactly: *c* causes *e* iff there is a (possibly empty) set of actual events, Σ, and set of actual omissions, Ω, such that if *c* were to occur without any of the events in Σ, and with all the omissions in Ω, then *e* might occur as a descendant of *c* with only actual events mediating between *c* and *e*. (Take an 'omission' to be a pair $\langle k, r \rangle$, where *k* is an event-kind and *r* is a spatiotemporal region, such that no event of kind *k* occurs at *r*. On omissions, see Lewis 1986b: 189–93.) But since the revised analysis offers a logically weaker condition for *c* to cause *e* than the original, this merely compounds the problem of insufficiency, to be discussed below.

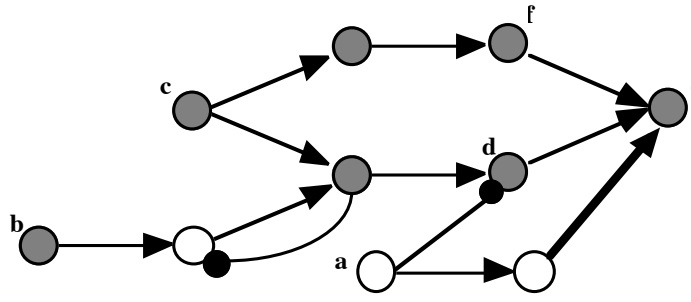

Figure 5

In this counterfactual situation, *e* occurs and is a descendant of *c* (had *c* not occurred, **e** would not have fired, for it would only have been stimulated by **d**.) The only non-actual events are *d*, and those in the causal chain leading from *d* to *e*. Thanks to the occurrence of *b*, these events are not descendants of *c* (had *c* not occurred, *b* would have brought them about instead). Therefore there are no *non-actual* events that mediate between *c* and *e*. According to the PCA-analysis, then, *c* causes *e*, and thus the analysis does not give a sufficient condition. (Again, no apparent problem for Lewis.)

Stating the 'only actual events mediating' clause in terms of *causal* mediation would solve the problem posed by figure 5: in that counterfactual situation causal intermediates (*d*, for example) occur which do not *actually* occur. The price, of course, is that the analysis is no longer reductive. But there is more to pay. For consider figure 6:

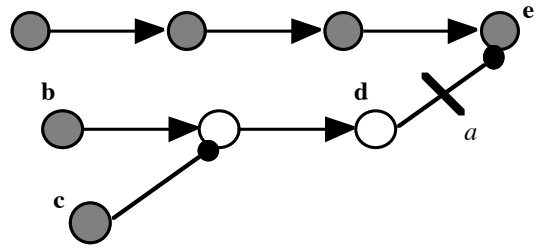

Figure 6

Here, the solid line between **d** and **e** indicates that some event *a* has occurred (long before **b** and **c** fire, let us suppose) which has *severed* the inhibitory connection between **d** and **e**. Thanks to the occurrence of *a*, *e* does not depend on *c*. More to the point, *c* is not a cause of *e*: with the inhibitory connection severed, the occurrence of *c* is wholly irrelevant to *e*. However, had *c* but not *a* occurred, the sequence of events would have been as depicted in figure 7:

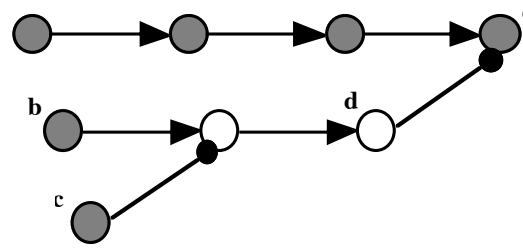

Figure 7

In this situation, does any event occur that is an effect of *c* and a cause of *e* – but which does not occur in the figure 6 situation? No: the events that causally mediate between *c* and *e* (what few of them there are!) are all *actual*.[5] Still, *e* is a descendant of *c*. Hence, the revised non-reductive PCA-analysis (and, indeed, the original) misclassifies *c* as a cause of *e*.[6,7]

We conclude that Lewis's second attempt remains the best candidate for a counterfactual analysis of causation.[8,9]

Massachusetts Institute of Technology
Cambridge MA 02139, USA
abyrne@mit.edu
ejhall@mit.edu

References

Ganeri, J., P. Noordhof, and M. Ramachandran. 1996. Counterfactuals and preemptive causation. *Analysis* 56: 216–25.

Hall, N. 1997. Two concepts of causation. Unpublished manuscript.

Lewis, D. 1973. Causation. *Journal of Philosophy* 70: 556-67. Reprinted in his 1986a, 159–72.

Lewis, D. 1986a. *Philosophical Papers*, Volume 2. New York: Oxford University Press.

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

[5] So, for that matter, are the omissions.

[6] A quicker counterexample to the sufficiency of the PCA-analysis, involving temporal action at a distance, is this. Imagine a faded arrow going from **a** to **e**, and a faded arrow going from **c** to **e**. The laws are such that **a**'s firing causes **e** to fire after $t+d$ (if **e** has not already fired), and **c**'s firing causes **e** to fire after t. **c** and **a** fire simultaneously at time 0, and **e** fires at time t. But only *c* causes *e*: the time delay between *a* and *e* is not long enough for *a* to cause *e*. Such a case is intuitively possible. Thanks to the lack of any intermediates, the PCA-analysis wrongly counts *a* as a cause of *e*. However, for present purposes more elaborate but less controversial examples are desirable.

[7] Consider the figure 7 situation. Lewis's account classifies *c* as a cause of *e*, we think correctly. But, interestingly, this situation involves a *kind* of temporal action at a distance.

[8] As argued in Hall 1997 this second attempt is ultimately unsuccessful.

[9] Thanks to Murali Ramachandran for discussion and for saving us from an error.