On Lewis, Schaffer and the Non-Reductive Evaluation of Counterfactuals

by

ROBERT NORTHCOTT

University of Missouri-St Louis

Abstract: Jonathan Schaffer (2004) proposes an ingenious amendment to David Lewis's semantics for counterfactuals. This amendment explicitly invokes the notion of causal independence, thus giving up Lewis's ambitions for a reductive counterfactual account of causation. But in return, it rescues Lewis's semantics from extant counterexamples.

I present a new counterexample that defeats even Schaffer's amendment. Further, I argue that a better approach would be to follow the causal modelling literature and evaluate counterfactuals via an explicit postulated causal structure. This alternative approach easily resolves the new counterexample, as well as all the previous ones. Up to now, its perceived drawback relative to Lewis's scheme has been its non-reductiveness. But since the same drawback applies equally to Schaffer's amended scheme, this becomes no longer a point of comparative disadvantage.

Keywords: counterfactuals, causation, reduction.

1. Introduction

JONATHAN SCHAFFER (2004) PROPOSES an ingenious amendment to David Lewis's semantics for counterfactuals. This amendment explicitly invokes the notion of causal independence, thus giving up Lewis's ambitions for a reductive counterfactual account of causation. But in return, it rescues Lewis's semantics from extant counterexamples.

In this article, I present a new counterexample that defeats even Schaffer's amendment. Further, I argue that a better approach would be to follow the causal modelling literature and evaluate counterfactuals via an explicit postulated causal structure. This alternative approach easily resolves the new counterexample, as well as all the previous ones. Up to now, its perceived drawback relative to Lewis's scheme has been its non-reductiveness. But since the same drawback applies equally to Schaffer's amended scheme, this becomes no longer a point of comparative disadvantage. I conclude that, since in all other respects the performance of the causal modelling approach is clearly preferable, we should endorse it over any non-reductive version of Lewis's scheme.[1]

---

1   Note that the critique of Lewis/Schaffer in this article will be entirely independent of Paul Noordhof's (2005) dispute with Schaffer.

2. The Case of Multiple Effects

Schaffer amends Lewis thus (2004, p. 305):

"If it were p, then q" is true at world w iff: if there are p-worlds, then there is a $p\&q$-world closer to w than any $p\&{\sim}q$-world.

Possible worlds are ordered by comparative similarity according to these criteria:

(1) It is of the first importance to avoid big miracles.

(2) It is of the second importance to maximize the region of perfect match, *from those regions causally independent of whether or not the antecedent obtains.*

(3) It is of the third importance to avoid small miracles.

(4) It is of the fourth importance to maximize the spatiotemporal region of approximate match, *from those regions causally independent of whether or not the antecedent obtains.*

The amendments are the added italicized sections in (2) and (4); Lewis's original (1979) similarity criteria are given by subtracting these sections.

Before presenting the new counterexample, it will be useful to begin with the following related case: *Smoking*. Let c = a man smoking in a poorly ventilated room, and let the following be three effects of c: e_1 = the smell of tobacco smoke gets caught in his hair; e_2 = his teeth become yellowed; and e_3 = his eyes itch.[2,3] The case is one of token causation.
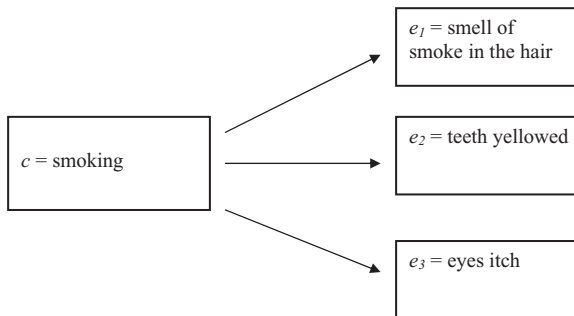

Figure 1 – Smoking

Now consider the counterfactual:

If $(\sim O(e_1)$ & $\sim O(e_2)$ & $\sim O(e_3))$, then $\sim O(c)$ (5)

---

2 A case with this structure is discussed by James Woodward (2003, pp. 139–141).

3 For ease of exposition, in the text I use c, e_1, etc. to denote actual events. But, strictly speaking, in the causal graphs they denote *variables*, and the events merely instantiate the actual values of those variables.

(Let $O(x)$ be the proposition that the event x occurs.) In words, if we were to clean his hair with shampoo, re-whiten his teeth with polish, and relieve his eyes with droplets, we would thereby prevent his smoking.[4] Intuitively, (5) is clearly false – just treating symptoms has no impact on the cause.

Label the actual world w_0. Lewis's criteria analyse the case thus: imagine a possible world w_1, featuring three miracles, one just before each e_i, and each negating that e_i. And imagine another world w_2, featuring only one negating miracle, this time just before c. (So in both w_1 and w_2, ($\sim e_1$ & $\sim e_2$ & $\sim e_3$).) Endorsement of w_1 as the closer to w_0 would yield the desirable answer that (5) is false since c would be left intact, whereas w_2 would yield the answer that it is true. But w_1 requires three miracles, w_2 only one. This makes w_1 reachable only via a "big" miracle.[5] True, w_1 does achieve a slightly greater region of perfect match, but by the priority of criterion (1) over (2) it seems we must judge w_2 to be the closer to w_0, and hence judge (5) to be true. (If this does not seem obvious, the number of common effects may be multiplied indefinitely until the comparison is overwhelming.)

What of Schaffer's amendments? Schaffer argues (and I agree) that they neatly resolve several existing counterexamples to Lewis's scheme. Can they similarly resolve Smoking? Unfortunately, no. The relevant regions of match are those between c and the e_i, i.e., regions causally upstream of the e_i. But Schaffer's amendments only rule out consideration of regions causally *downstream* of the e_i, so the Lewisian analysis is unaltered here. Moreover, in any case the amendments still prioritize counting up miracles. Thus we still end up with the wrong answer for (5).

One response to examples such as Smoking is that counterfactuals need to be defined in terms of (the occurrence of) *events*. In particular, because of its complex spatially separated nature, ($\sim e_1$ & $\sim e_2$ & $\sim e_3$) does not count as a *bona fide* event and thus, so the argument runs, there is no obligation on Lewis's scheme to evaluate counterfactuals with such an antecedent.

In the next section, I present a revised version of the example that addresses this objection. Before that, though, it is worth questioning how compelling the objection really is anyway. As a practical matter, we are very much interested in counterfactuals with complex antecedents. For instance, they are essential for separating

---

4   We are not necessarily committing ourselves to negative events here. For instance, the negations might be read as shorthand for contextually salient alternative positive events.

5   The distinction between big and little miracles is left by Lewis a little vague. But it seems clear here that ($\sim e_1$ & $\sim e_2$ & $\sim e_3$) would indeed qualify as a bigger miracle than $\sim c$: "We can . . . distinguish one simple unlawful event from many, or from one complex event consisting of many simple unlawful parts. A big miracle consists of many little miracles together, preferably not all alike. What makes the big miracle more of a miracle is . . . that it is divisible into many and varied parts, any one of which is on a par with the little miracle" (Lewis, 1986a, p. 56).

a cause's direct and indirect impacts on an effect (e.g., Woodward, 2003, pp. 45–61). They are ubiquitous in science, and indeed everyday life. For example, would washing your hair, polishing your teeth and so on together be a good way to prevent smoking? It seems a significant limitation of scope to be unable to address such questions – and a limitation not shared, as we will see shortly, by the causal modelling literature.

Moreover, Lewis himself came to argue that, at least in absence cases, we may evaluate counterfactuals without requiring events as relata (e.g., Lewis, 2004, pp. 282–283). So why then insist on (simple) events in Smoking?[6]

3. A New Counterexample

Consider an augmentation of Smoking, label it *Salon*. To remedy the smelly hair, yellow teeth and itchy eyes, suppose the smoker takes a trip to the salon (event *s*). That is, each of these three cosmetic problems independently causes *s*.[7]
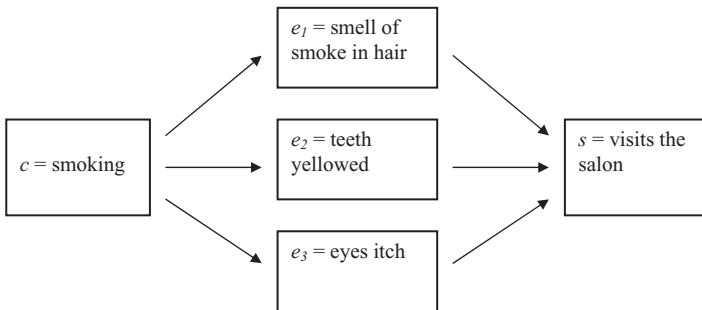

Figure 2 – Salon

Now consider the counterfactual:

If ~O(*s*), then ~O(*c*)   (6)

In words, if we were to prevent the smoker going to the salon afterwards, we would thereby also prevent the smoking. Again, intuitively (6) is clearly false, since *s* is an

---

6   Further, as I read him Lewis is also uncertain that entities such as (e_1 & e_2 & e_3) need even be disallowed as events in the first place: "I leave open the question whether several events, however miscellaneous, always have another event as their sum . . . it seems hard to tell when we can be content to say only that several events are joint causes and separate effects, and when we must also insist on a single event that is their sum" (1986b, p. 260).
7   This addition to the story was suggested to me by Juan Montana.

effect of c, not its cause. And this time, the antecedent is undoubtedly a kosher single event.

Let us apply Lewis's criteria. Given its overdetermination, it apparently requires *three* miracles to prevent s while leaving c intact. In other words, it requires a single big miracle. In contrast, still only a single small miracle is required to prevent s by preventing c. Therefore the nearest $\sim s\&\sim c$-world must be deemed closer to actuality than any $\sim s\&c$-world. Accordingly, (6) comes out true, the opposite of what is desired. Once again, Schaffer's amended criteria fare no better: the relevant region of match is between c and s, all of which is causally upstream of the antecedent $\sim s$, and so we are left with the same incorrect final result.

But, it might be objected, how do we know that the context does demand consideration of a big miracle? Might not the *nearest* $\sim s$-world rather be one reached via just a small miracle, say the salon closing or the smoker forgetting to go? In which case, as desired, the Lewis (and Schaffer) scheme could declare (6) false after all.

Two replies:[8] first, I see no way to guarantee that the relative size of the relevant miracles always turns out right. Or at least, I see no way to guarantee this save by illicitly smuggling pre-existing causal intuitions into our interpretation of the similarity criteria. For example, suppose our smoker's character is such that he often absent-mindedly forgets to light a waiting cigarette, but by contrast he is always steadfast in keeping salon appointments. In such a context, achieving $\sim c$ may require only a single neuron not to fire, but (given c) achieving $\sim s$ would require something rather more dramatic – either many neurons firing differently, or perhaps instead some fire or public holiday to close the store, or indeed the three separate miracles originally suggested. So now the nearest $\sim s\&c$-world would be a big miracle away, the nearest $\sim s\&\sim c$-world only a small one. I do not see how it is possible to rule out such gerrymandered contexts. Yet in any of them, Lewis's scheme must declare (6) true – even while it remains obstinately false.

A second reply: as Woodward notes (2003, p. 141), we may in any case tweak the original story by making the occurrence of c *chancy*, perhaps indeed very unlikely.[9] In Salon, that would mean $\sim c$ not requiring any miracle (or quasi-miracle) at all, whereas (given c) $\sim s$ still does. An incorrect evaluation of (6) by Lewis's – and Schaffer's – criteria then appears unavoidable.

Another objection to Salon might run: is it merely trading on the well-known problem of overdetermination, common to all counterfactual theories of causation? No – generally in overdetermination puzzles, there is no controversy over the

---

8   Each applies to cause-effect pairs generally, not just to cases with the structure of Salon.

9   The arguments of this article apply equally to indeterministic cases – indeed more so, since while causal models handle them straightforwardly, Lewis must invoke "quasi-miracles" which raise independent difficulties of their own (examples of which Schaffer himself notes, 2004, p. 304). See also the next section.

counterfactual dependencies. Rather, it is acknowledged by all that the effect is not counterfactually dependent on any of the causes individually, and the difficulties arise only once we invoke counterfactual dependence as a criterion for causation. But in Salon, the dispute is over that initial judgement of counterfactual dependence itself. Moreover, the counterfactual dependence in question is not that between the effect and one of its overdetermining causes, i.e., not that between s and one of e_1, e_2 or e_3. Rather, it is the dependence between s and c.

4. A Better Way?

Turn now to the burgeoning recent literature on causal modelling (e.g., Pearl, 2000; Spirtes *et al.*, 2000). This does not evaluate counterfactuals in the Lewisian manner, i.e., does not evaluate them by judging similarity between possible worlds according to nomological considerations. Instead, it evaluates them on the basis of a causal model, i.e., explicit structural equations often in combination with a causally interpreted graph. For example, Figure 1 might be interpreted as a causal graph with appropriate associated equations, and (5) evaluated by tracing the impacts of *interventions* in that graph. Roughly speaking, an intervention changes the value of a particular variable in isolation and we then use the equations to calculate the impacts of that change on other variables. (I do not discuss here the details of how interventions should be defined.)

In Smoking, that would mean intervening on each of the e_i individually and then calculating the resultant impact on c, which the equations and graph enable us to do immediately. In particular, since each e_i is causally downstream of c, these interventions would have no effect on c. Therefore, just as intuition demands, ($\sim e_1$ & $\sim e_2$ & $\sim e_3$) would *not* imply $\sim c$ and so (5) comes out false. Similarly, in Salon, the counterfactual antecedent $\sim s$ corresponds in Figure 2 to an intervention on s and thus, since this is causally downstream of c, would again imply that c is unaffected and so that (6) is false.

A causal modelling approach also easily handles other counterexamples to Lewis's scheme, in particular those which Schaffer's amendment is designed to overcome. Take *Morgenbesser's coin* as a representative example (Schaffer, 2004, p. 303): consider an indeterministic coin flip. While the coin is in mid-air, Fred bets heads. The coin lands tails, so Fred loses. The following counterfactual seems intuitively true:

If Fred had bet tails, then he would have won   (7)

Let b = Fred makes his bet (b taking the value either heads or tails), f = the outcome of the coin flip (again either heads or tails), and o = the outcome of the bet (either

Fred wins or loses, depending on the values of *b* and *f*). Then the following causal graph (plus appropriate structural equations) would capture the example:
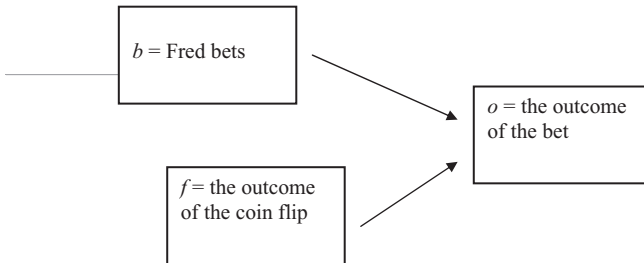

Figure 3 – Morgenbesser's Coin

In Figure 3, an intervention on Fred's bet, changing it from heads to tails, will clearly have no impact on the outcome of the coin flip, since *f* is not causally downstream of *b*. Thus Fred would now win the bet, and so (7) comes out true, as desired.[10]

The ability to handle such examples successfully is unsurprising. The notion of an intervention is itself a causal notion, and a causal model leaves us free to insert such interventions exactly where intuition demands. In particular, interventions can be freely made directly on the e_i or s. Lewis's scheme, by contrast, crucially fails to build in any preference for these targeted interventions in such cases, achieving them – when it does – only indirectly via the totting up of miracles and matches of fact. This of course is the price paid for a reductive account, but the very same unsatisfactorily indirect method is now inherited by Schaffer's amended scheme even though the associated reductive aspirations have been abandoned. And it is the reliance on acausal similarity criteria rather than the causal notion of intervention, that leads Schaffer's scheme astray in Salon.[11,12]

---

10   The graph also explains immediately why if, rather than change his bet, Fred had instead swatted the coin in mid-air – thereby re-randomizing the outcome of the flip – we would now intuit merely that he *might* have won his bet, not that he definitely would have (Schaffer, 2004, p. 303 n. 10). This new situation would correspond to adding a new box to represent Fred's swatting, that box being connected to *f* by a (probabilistic) causal arrow.

11   Adopting a causal modelling approach still leaves us considerable flexibility regarding our ultimate ontological commitments. One possibility is that advocated in Schaffer (2004), namely to regard causal and counterfactual facts as "co-supervenient" upon a Humean base. (Thus there is no need to stick to a Lewisian semantics for that.) Another possibility is a non-Humean ontology of causal powers. In addition, there is also already considerable debate over whether causation can be metaphysically identified with the formal structure of interventions (see, e.g., Woodward, 2003 and Cartwright, 2007 for contrasting views).

12   Arguably, explicitly contrastive theories of causation might serve a similar function here, i.e., of pinpointing exactly where we want an intervention to be (Schaffer, 2005; Northcott, 2008). Indeed, a contrastive view is implied by a causal modelling approach (Woodward, 2003).

Evaluation of counterfactuals via causal models carries other advantages too. For instance, there is none of the notorious vagueness surrounding Lewis's criteria, as each intervention can be defined explicitly and its consequences then traced precisely. Moreover, given certain assumptions, the machinery of causal equations and graphs can be of great practical use for inferring causal relations from acausal probabilistic and statistical dependence relations. (Indeed, there is now a mini-industry exploiting such Bayes-net methods.)

5. Conclusion

Compared to Lewis's account, the philosophical Achilles heel of the causal modelling approach has been its non-reductiveness. But if Schaffer is right that we must give up Lewis's reductive ambitions in order to make his similarity criteria resistant to counterexamples, it follows that this comparative disadvantage disappears. Given its many other advantages, the causal modelling approach is then clearly preferable to any non-reductive version of Lewis's semantics for counterfactuals. Lewis's scheme, if it is to be saved at all, must be saved some other way.

References

CARTWRIGHT, N. (2007) *Hunting Causes and Using Them: Approaches in Philosophy and Economics*. New York: Cambridge University Press.

LEWIS, D. (1979) "Counterfactual Dependence and Time's Arrow." *Nous*, 13: 455–476.

LEWIS, D. (1986a) "Postscripts to 'Counterfactual Dependence and Time's Arrow'." In D. Lewis, *Philosophical Papers: Volume II*, pp. 52–66. Oxford: Oxford University Press.

LEWIS, D. (1986b). "Events." In D. Lewis, *Philosophical Papers: Volume II*, pp. 241–269. Oxford: Oxford University Press.

LEWIS, D. (2004) "Void and Object." In J. Collins, N. Hall and L. A. Paul (eds) *Causation and Counterfactuals*, pp. 277–290. Cambridge, MA: MIT Press.

NOORDHOF, P. (2005) "Morgenbesser's Coin, Counterfactuals and Independence." *Analysis*, 65: 261–263.

NORTHCOTT, R. (2008) "Causation and Contrast Classes." *Philosophical Studies*, 39, 111–123.

PEARL, J. (2000) *Causality*. New York: Cambridge University Press.

SCHAFFER, J. (2004) "Counterfactuals, Causal Independence, and Conceptual Circularity." *Analysis*, 64: 299–309.

SCHAFFER, J. (2005) "Contrastive Causation." *Philosophical Review*, 114(3): 297–328.

SPIRTES, P., GLYMOUR, C. and SCHEINES, R. (2000) *Causation, Prediction, and Search* (second edition). Cambridge, MA: MIT Press.

WOODWARD, J. (2003) *Making Things Happen: A Theory of Causal Explanation*. New York: Oxford University Press.