

Counterfactual Dependence and Time's Arrow

20 November 2006

1. The context-sensitivity of counterfactuals

Backtracking versus non-backtracking contexts.

In non-backtracking contexts, we seem to feel very free to use auxiliary premises about times before t in figuring out what would have been the case if matters had gone differently at t . EG: 'if I had run a four-minute mile yesterday, it would have been a first for me'.

2. A paradox about counterfactuals under determinism.

Suppose determinism is true. Then the propositions that are laws of nature L , and the complete description I of conditions very very long ago, jointly entail that I come to class today. Our paradoxical argument:

- (1) If I hadn't come to class today, L would have still been true.
- (2) If I hadn't come to class today, I would have still been true.
- (3) If I hadn't come to class today, ' $(L \wedge I) \supset I$ come to class today' would have still been true.
 - Argument: true counterfactuals with metaphysically possible antecedents always have metaphysically possible consequents.
- (4) Therefore, if I hadn't come to class today, ' $(L \wedge I \wedge ((L \wedge I) \supset I \text{ come to class today}))$ ' would have still been true.
 - Argument: Agglomeration is valid.
- (5) Therefore, if I hadn't come to class today, I would have come to class today.
 - Argument: counterfactuals (at least those with logically consistent antecedents) are closed under logical consequence in the consequent.

5 seems obviously false—and it seems that we know much better that it is false than that determinism is false! Main options for resisting the argument: deny 1 (Lewis); or deny 2.

3. Denying (2) while retaining (1)

If the (2)-denier wants our ordinary evaluations of non-backtracking counterfactuals to come out generally true, the thought will have to be this: if I hadn't come to class today, things would have been different all the way back; but until very recently, they would have been only *very slightly* different, to such an extent that almost all ordinary claims about the past would still have been true.

- Is this enough?

Lewis:

‘There is no guarantee whatever that [a possible world] can be chosen so that the differences diminish and eventually become negligible in the more and more remote past. Indeed it is hard to imagine how two deterministic worlds anything like ours could possibly remain just a little bit different for very long. (p. 45)’

But surely the *continuity* of the laws of nature is a good reason to think that for any t , we can find possible initial conditions so close to the actual initial conditions that the world they generate is only negligibly different from the actual world until t .

Moreover, *chaos* theory suggests that for a great many ordinary processes, small differences blow up *very* rapidly into big differences. Good news for (2)-deniers!

4. Denying (1): miracles

Lewis’s picture: when A is false and entirely concerns what’s going on at t , what would have been the case if A is this: things would have been just as they actually are until *shortly before* t ; then there would have been a small miracle (an exception to one of the generalisations that are actually laws of nature), of just the right sort to lead to its being the case that A ; after that, there would have been no more miracles.

5. Asymmetry by fiat.

6. Lewis’s criteria for closeness.

- (1) It is of the first importance to avoid big, widespread, diverse violations of law.
- (2) It is of the second importance to maximize the spatio-temporal region throughout which perfect match of particular fact prevails.
- (3) It is of the third importance to avoid even small, localized, simple violation of law.
- (4) It is of little or no importance to secure approximate similarity of particular fact, even in matters that concern us greatly.

Motivation: we want $w1$ —a world exactly similar to the actual world until just before t , at which point there is a small miracle which makes Nixon press the button, after which things proceed in accordance with the actual laws, so that a nuclear holocaust ensues—to come out closer to the actual world ($w0$) than any of the following competitors (in all of which Nixon presses the button).

- $w2$: a world where the *entire* past is different and there are no miracles at all.
- $w3$: a world where the past up to shortly before t is the same, whereupon a small miracle leads to Nixon’s pushing the button, after which another small miracle leads to the signal’s not getting all the way to the launcher.

*Crucial claim: there's no way for a *small* miracle to bring about *perfect* reconvergence: so any w_3 -type world won't *perfectly* match the actual world at later times.

- w_4 : a world where the past up to shortly before t is the same, whereupon a small miracle leads to Nixon's pushing the button, whereupon a *big* miracle wipes out all of the traces of his having done so (fingerprints, click on tape, light signals shooting out into space...)

Lewis's plausible claim: there are no possible worlds like ours except for *two* small miracles, that exactly match our world except during the period between those two miracles. [NB: this is temporally symmetric]. Lewis seems to confuse this with:

Lewis's implausible claim: Our world is 'a world to which convergence is difficult'. There are no possible worlds like the actual world but for one small miracle, which match the actual world perfectly after the miracle but not beforehand (and in which, e.g., Nixon presses the button).

7. Elga's objection

"Lewis's implausible claim" is false. Trick to seeing this: play the history of the world in reverse. Note how perfectly everything seems to conspire to produce entropy-decreasing options. These processes are very fragile: a tiny miracle should be enough over time to disrupt everything and set the world [viewed in reverse] on a more normal, entropy-increasing trajectory.

So there are worlds containing one small miracle shortly after t , that are exactly like the actual world after t , but are quite radically different before that: for most of the time before that, they look like the actual world but played in reverse.

- Problem: arguably, actual people like Nixon don't exist at any such worlds, so none of them is a world where Nixon presses the button. But this won't help us when we're dealing with counterfactuals that have negative antecedents.

Possible fix: adopt an Albert-style view of statistical mechanics. There's a Past Hypothesis—some claim about the distant past that entails that it's got very low entropy—that's a law. Elga's worlds violate this law.

8. A paradox about counterfactuals under widespread indeterminism

Setup: I steal a fair coin that you were about to toss a million times. Let $S_1...S_{2^{1,000,000}}$ be all the possible sequences of a million heads and tails.

Consider the following counterfactuals:

(C_n) If I hadn't stolen your coin, you wouldn't have got sequence S_n .

Which of these is true?

Answer 1: none of them.

- Problem: entails that if indeterminism is widespread, our ordinary counterfactual judgments are almost always false.

Answer 2: all of them.

- Problem: violates Agglomeration.

Answer 3 (Lewis): all the ones where S_n is not 'remarkable'

- Problem: it seems awful to say 'Although if I hadn't stolen your coin S_1 and S_n would have been equally likely, and it's not the case that if I hadn't stolen your coin you wouldn't have got S_n , it is the case that if I hadn't stolen your coin you wouldn't have got S_1 .
- Further problem: how to generalise so as to yield intuitive evaluations of counterfactuals at worlds with widespread indeterminism?

Answer 4 (Stalnaker?): one is true and the rest are false, but it's indeterminate which.

- Problem: entails that if indeterminism is widespread, our ordinary counterfactual judgments are almost never determinately true.

Answer 5 (Hawthorne): one is true and the rest are false, and we should have the same low degree of belief in each.

- Problem: conflicts with Humeanism.
- Unless we somehow combine it with an expressivist semantics?...

9. Expressivism to the rescue?