

The Asymmetry of Counterfactual Dependence

PHIL 1480 Metaphysics

Cian Dorr

30 March 2006

1. Temporal asymmetries

Ways in which the future direction is unlike the past direction — i.e. ways in which the 'later than' relation is unlike the 'earlier than' relation.

- The asymmetry of causation: earlier events cause later ones, not v.v.
- The asymmetry of knowledge: at any time, we can know much more about how things were before that time than we can know about how things are after that time.
- The obscure asymmetry of 'openness'.
- The asymmetry of counterfactual dependence:

2. Counterfactual (or subjunctive) conditionals

'If Oswald hadn't killed Kennedy, no-one would have.'

- Contrast with the *indicative* conditional 'If Oswald didn't kill Kennedy, no-one did'.

Abbreviate 'If it were the case that P, then it would be the case that Q' as 'P → Q'.

The asymmetry: if things had been different at *t*, then things would have been different after *t* in all sorts of ways. But our standard way of reasoning about counterfactuals seems to commit us to saying that things would have been just the same, or at least almost the same, at times before *t*.

The challenge: explain why this should be so, by giving an analysis of what counterfactual conditionals mean.

3. Warmup puzzle: counterfactuals under determinism

Let *determinism* be the following claim: the laws of nature, together with the facts about any stretch of history, entail all the facts about all of history. Suppose it's true. What would have been the case if I had not come to class today?

Two answers are possible:

- (i) the entire past would have been different—even the very distant past.
- (ii) the actual laws of nature would have been violated. (Lewis's view.)

4. Asymmetry by fiat

One possibility is to *build a differential treatment of past and future right into the analysis*. This is the approach adopted by what Lewis calls 'Analysis 1'.

ANALYSIS 1. Consider a counterfactual “If it were that A, then it would be that C”, where A is entirely about affairs in a stretch of time t_A . Consider all those possible worlds w such that:

- (1) A is true at w ;
- (2) w is exactly like our actual world at all times before a transition period beginning shortly before t_A ;
- (3) w conforms to the actual laws of nature at all times after t_A , and
- (4) during t_A and the preceding transition period, w differs no more from our actual world than it must to permit A to hold.

The counterfactual is true if and only if C holds at every such world w .

Lewis’s complaints about this analysis:

- It doesn’t tell us what to say about counterfactuals whose antecedents are not entirely about affairs in a stretch of time.
- It incorrectly rules out backwards counterfactual dependence (outside the ‘transition period’) as absolutely impossible.

A further possible complaint: we now urgently need an explanation of why we have *this* concept rather than a different concept in which the roles of ‘past’ and ‘future’ are reversed.

5. Lewis’s strategy

Uses the philosophically useful vocabulary of *possible worlds*.

ANALYSIS 2. A counterfactual ‘If it were that A, then it would be that C’ is ... true if and only if some ... world where both A and C are true is more similar to our actual world, overall [**and in the relevant way**] than is any world where A is true but C is false.

- Under the circumstances we’ll be concerned with, this boils down to the same thing as: the A-world that is most similar to the actual world is a C-world.

What is the relevant ‘similarity’ (aka. ‘closeness’) relation? Here’s what Lewis says:

- (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.

Lewis’s claim: Assume for the sake of the example that the actual world is deterministic, and that Nixon’s button is wired straight up, at t , to the nuclear missile launcher. Then,

by this measure of similarity, the most similar worlds to the actual world where Nixon presses the button at t are ones which match actuality perfectly up to shortly up to t , then diverge via a small miracle (i.e. a violation of the *actual* laws of nature), then proceed according to the actual laws of nature, so that the nuclear holocaust ensues.

Let w_1 be such a world. What motivates Lewis's analysis of the relevant similarity relation is the need to get w_1 to come out closer to the actual world (w_0) than any of the following competitors, in all of which Nixon presses the button.

- w_2 —a world where the entire past is different and there are no miracles at all.
- w_3 —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.
- 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...)

6. Whence the asymmetry

The following temporally asymmetric fact is supposed to be true of our world: it takes only a small miracle to diverge from it, but it takes a big miracle to converge to it. I.e. any worlds that are exactly like it at all times after a certain time t but not beforehand contain a large, widespread violation of the laws of the actual world.

Is this true? Elga argues not.