

Counterfactuals

- Goodman's Account of Counterfactuals
- Two Problems with Goodman's Account
- Stalnaker's Account
- Lewis' Criticism of Stalnaker's Account
- Lewis' Account
- Two Aspects of Lewis' Account (Symmetry & Transitivity)
- Ramifications for Lewis' Theory of Causation

Goodman's Account of Counterfactuals

- Counterfactuals involve what would happen, or would have happened in specified hypothetical situations. For instance, one might say of a (unstruck, unlit) match, "had it been struck, it would have lit."
- Counterfactuals are called *counterfactuals*, because it is assumed that their antecedents did not, in fact, occur. As such, counterfactuals are not truth-functional. And, intuitively, they involve some kind of *modality*.
- Goodman's (1946) classic essay "The Problem of Counterfactual Conditionals" proposes the following analysis of the counterfactual conditional " $p \Box \rightarrow q$ " read "If p were true, then q would be true."
- $p \Box \rightarrow q$ if and only if p nomologically requires, given prevailing conditions, the truth of q . Or, in other words, if p , conjoined with some set of facts S and laws of nature L , deductively entails q . More formally, we have:

$$(G) \quad p \Box \rightarrow q \text{ if and only if } \{p, S, L\} \models q$$

Problem #1 for Goodman's Account: The Cotenability Problem

- Suppose (as is typically the case) that p and q are (in fact) both false. If we were to include $\neg q$ in the "set of facts" S , then Goodman's $\{p, S, L\}$ would be *unsatisfiable*, and so $p \Box \rightarrow q$ would be true, *trivially*.
- For example, we accept "If the match had been struck, it would have lit" because the match was dry and oxygen was present. We do not, however, (*pace* Goodman) accept "If the match had been struck, there would have been no oxygen" even though the match was dry and did not in fact light.
- Not wanting to make all (interesting) counterfactuals true, Goodman suggested that we constrain S in such a way that it include only facts "that would not be altered by the truth of (*i.e.*, those cotenable with) p ."
- The problem with this response is that it constrains S using, essentially, another counterfactual. In essence, Goodman's "fix" is to add the constraint that $p \Box \rightarrow S$. But, this is viciously circular.

Problem #2 for Goodman's Account: The Problem of Lawlikeness

- Goodman's account of counterfactuals (like our neo-Humean accounts of causation) invokes "laws of nature" L as an essential part of the analysis. What are laws? How do we distinguish them from non-laws?
- Consider the following two generalizations:
 1. All signals travel slower than light.
 2. All gold objects have mass less than 100,000 Kg.
- It is generally conceded that (1) is a law of nature (if true), and (2) is not (even if true). First, note that they are both of the same syntactical form, so it can't be anything formal that is making the difference.
- Goodman suggests a generalization of the form "All A 's are B 's" is a law if it is "projectible" — that is, if the observation of A 's that are B provides reason to believe that unobserved A 's are B .
- But, when does this happen? It seems that there are many non-lawlike

generalizations that are (legitimately) “projected” from small samples. Indeed, (2) seems to be an example. The high credibility of (2) does not seem to depend on the observation of every single gold object.

- All it takes for a generalization to be projectible is the belief that its truth would not just be a coincidence, but there is some uniform reason for its conformity — they cannot be “wholly accidental.”
- Hempel suggested that (2) be ruled-out on the grounds that if it were taken to be a law, then certain phenomena that our current theories to be perfectly possible would be precluded. Too strong? No new laws?
- This is an age-old problem. As with causation, there are Humean and non-Humean conceptions of laws. Humean conceptions of laws are based on regularities. Non-Humean conceptions posit (non-verbal) nomological necessity which outstrips mere regularity. A (or some) Humean would take (2) to be lawlike. But, non-Humeans would not. They would say it is nomologically possible for gold objects to be greater than 100,000 Kg. What if we replace “gold” in (2) with “uranium”? Do things change?

Stalnaker’s Account of Counterfactuals

- In his (1968) “A Theory of Conditionals”, Robert Stalnaker proposes a possible world semantics for counterfactual conditionals.
- Digression on possible world semantics for modal logics: “ $\Box p$ ” (necessarily, p) is interpreted as “ p is true in all worlds which are accessible from the actual world.” The accessibility relation R between worlds can have various properties. If R is transitive, then $\Box p \rightarrow \Box \Box p$ is valid. If R is symmetric, then $\models \Box p \rightarrow p$. If R is serial, then $\models p \rightarrow \Box \Diamond p$, where “ $\Diamond p$ ” (possibly, p) is true if and only if “ $\neg \Box \neg p$ ” is true.
- Stalnaker introduces a selection function $f(p) = w$ which maps each false sentence (or proposition) p into “the world w such that w is the minimal revision of the actual world that would be required to make p true.”
- Thus, we have the following Stalnakerian account of counterfactuals:

$$(S) \quad p \Box \rightarrow q \text{ if and only if } q \text{ is true at } f(p)$$

Some Features of Stalnaker’s Account

- Stalnaker’s account has various virtues. It is able to match certain intuitions about non-truth-functional conditionals. For instance:

If I put sugar in this cup of tea it will taste fine.
 \therefore If I put sugar and diesel fuel in this cup of tea it will taste fine.

is not a valid argument on Stalnaker’s account. This is as it should be.
- Stalnaker’s assumption that there is a *unique* closest p -world to the actual world turns out to be equivalent to the following pair:

$$(CEM) \quad (p \Box \rightarrow q) \vee (p \Box \rightarrow \neg q)$$

$$(SA) \quad \text{If } \Diamond p, \text{ then } \neg(p \Box \rightarrow q) \text{ iff } p \Box \rightarrow \neg q$$
- “CEM” stands for “Conditional Excluded Middle” and “SA” stands for “Stalnaker’s Axiom”. It is these assumptions that Lewis takes as a point of departure in his (similar) theory of counterfactuals.

Lewis’ Objections to Stalnaker’s Account

- In his (1972) paper “Counterfactuals and Comparative Possibility”, David Lewis takes Stalnaker to task for his account of counterfactuals. In particular, Lewis objects to the assumption that there is a *unique* $f(p)$.
- Lewis argues that for an antecedent like “Bizet and Verdi are compatriots”, a world in which they are both French and a world in which they are both Italian may be equally close to the actual world, and both might be closer than any others. In such a case, both:

It is not the case that if Bizet and Verdi were compatriots Bizet would be Italian.
and $[i.e., \text{ both } \neg(p \Box \rightarrow q) \text{ and } \neg(p \Box \rightarrow \neg q)]$
It is not the case that if Bizet and Verdi were compatriots Bizet would not be Italian.
could come out true on Lewis’ analysis, directly contradicting (CEM).
- If there is a unique closest p -world, then Lewis’ analysis agrees with Stalnaker’s. If there are ties, however, the two accounts will diverge.

Lewis' Account of Counterfactuals

- Lewis proposes the following variation on Stalnaker's approach:
 $(\mathcal{L}) \ p \Box \rightarrow q$ if and only if some (accessible) $p \ \& \ q$ -world is closer to the actual world than any $p \ \& \ \neg q$ -world, if there are any accessible p -worlds. If there are no accessible p -worlds, then $p \Box \rightarrow q$ is true.
- Lewis rejects two assumptions implicit in Stalnaker's account.
 (Uniqueness Assumption) the assumption that there is a unique closest p -world to the actual world, and (Limit Assumption) as we proceed to closer and closer p -worlds, we eventually hit a limit and can go no farther.
- Lewis and Stalnaker both accept the following two assumptions:
 (Ordering Assumption) comparative similarity imposes a weak ordering (connected and transitive) on the accessible worlds, (Centering Assumption) the actual world is accessible from itself (indeed, they assume reflexivity of R), and closer to itself than any other world is to it.

Two Aspects of Lewis' Account of Counterfactuals I: Asymmetry

- Counterfactuals are not anti-symmetric. That is, it is quite possible to have both $p \Box \rightarrow q$ and $q \Box \rightarrow p$, even if p and q are logically independent.
- Lewis tells an elaborate story (in his "Counterfactual Dependence and Time's Arrow") about how to make judgments of relative similarity in cases involving real events. This is supposed to provide for "natural" asymmetries, without appealing explicitly to temporal asymmetries.
- We are told that the most similar worlds are those in which our laws of nature (laws again!) are rarely violated. But, exact similarity with respect to facts in some large region of spacetime is also a major factor and will promote similarity even at the cost of minor "miracles".
- Some have objected to Lewis' theory on the grounds that it seems to say that "If the President had pressed the button, a nuclear war would have ensued" is false, since a world in which the circuit fails and war is averted is more similar to the actual world than a world with destruction.

- Lewis' reply is that the truth of the conditional should tell us that we've got our judgment of similarity wrong, and that we need to mind the following importance ranking of world differences: (1) existence of many miracles, (2) absence of an exact matching of particular facts over large regions of spacetime, (3) existence of a small number of miracles.
- Worry: This ordering may give the right answers, but where does it come from other than by reverse engineering the true conditionals? Let's look more carefully at Lewis' "asymmetry of overdetermination" argument.
- Imagine a hypothetical change in the course of our world — that some event p at time t did not occur. Lewis claims it would be hard to reconcile this supposition with what actually happened after t , since p brought about many phenomena which determine that p did occur at t .
- But, Lewis claims, it is relatively easy to square the supposition that p did not happen with the course of the world before t . This is supposed to be because events are not substantially overdetermined by what preceded them. So, at the cost of a 'small miracle', we can reconcile the

nonoccurrence of p with the actual history of the world before t .

- However, the argument continues, we cannot, without much greater cost, reconcile this with the actual future of the world after t . Consequently, among possible worlds without p , those that are just like ours until t and then diverge are more similar to the actual world than those that are just like ours after t or those that differ from ours before t .
- Thus, from the "asymmetry of overdetermination" it follows that that if the present were different from the way it is, then so the future would be different, but not the past. Hence, there will normally be no chain of counterfactually dependent events leading backward in time. Voil!
- There are two keys to this argument. First, there is the "asymmetry of overdetermination" assumption, that events are more strongly overdetermined by their futures than by their pasts. Second, there is the three-tiered "importance ranking of world differences".
- Indeed, these work in concert to bring about the asymmetry Lewis wants. But, why should we believe either of the key assumptions?

Two Aspects of Lewis' Account of Counterfactuals II: Transitivity

- Lewis' counterfactuals are not transitive either. That is, we can easily have $p \Box \rightarrow q$, $q \Box \rightarrow r$, and $\neg(p \Box \rightarrow r)$. This explains why Lewis takes causation to be the *ancestral* of counterfactual dependence.
- Traditionally, people have taken (deterministic) causation to be asymmetric (temporally, in fact), transitive, and irreflexive. Lewis goes to great lengths to try to get (temporal) symmetry into his account. He also makes sure to get transitivity (by fiat). This he gives no argument for, really [see Ned Hall's paper, in next week's readings]. How does he get irreflexivity? Or, does he get this into his account of causation?
- In other areas (*e.g.*, preference theory), counterexamples to transitivity are often generated from contextual considerations. How could Lewis change his account of counterfactuals to allow for contextuality? What might a counterfactual account of " p caused q in C " look like?

Some Basic Worries About Counterfactuals and Causation

- There are some basic (and somewhat silly) worries about counterfactuals and causation. First, there is the reflexivity problem (previous slide). For all p , we have $p \Box \rightarrow p$. Not so for causation.
- This is a special (trivial) case of a more general phenomenon: that counterfactual dependence can be non-causal. Other examples include:
 - If John had not been killed, his wife would not have been a widow.
- How does Lewis deal with these cases?
- What do you think of the following counterfactual example? You read in the paper that the winning lottery number is 12. You say to yourself, "Had I bought a lottery ticket with the number 12, I would have won the lottery." Is this true on Lewis' account? Should it be true? What does this say about the relationship between counterfactuals (when they are properly understood) and causation (intuitively)?