Conditionals, Part II: Philosophical Logic.

Lecture V, *More of Lewis on Counterfactuals*, 7th November.

Christopher J. Masterman (cm789@cam.ac.uk, christophermasterman.com)

Last week, we saw Lewis's argument for why $\Box\!\!\rightarrow$ cannot be identified straightforwardly as a strict conditional. Rather, it should be identified with a *variably* strict conditional. The resulting account is similar in many respects to Stalnaker's semantics, but there are some significant differences.

1. *Lewis's Analysis of Counterfactuals*

1.1. Begin with the intuitive idea. For each possible world w, we can construct a system of spheres containing other possible worlds, emanating out from w. Every world within a particular sphere is equally similar (similar to at least some degree) to w. The smaller the sphere around w, the more similar the worlds in that sphere are to w. We then give our semantics in terms of these systems of spheres.

1.2. In more detail, first let \$ be an assignment to each possible world i of a set $\$_i$ of sets of possible worlds. Each of these sets $\mathcal{S} \in \$_i$ we call *spheres*. We focus on a particular kind of set of spheres—a *centred system of spheres*. Without too much formal detail: the centred system $\$_i$ is a *nested* series of spheres, with i at the centre and it is closed under union and intersection. The truth conditions are given:

(*) $\phi \Box\!\!\rightarrow \psi$ is true at a world i, according to a system of sphere $\$_i$ iff either

  (1) no ϕ-world belongs to any sphere $\mathcal{S}$ in $\$_i$, or

  (2) some sphere $\mathcal{S}$ in $\$_i$ does contain at least one ϕ-world, and $\phi \supset \psi$ holds at every world in $\mathcal{S}$.

Note that there are two ways in which $\phi \Box\!\!\rightarrow \psi$ can fail to be true at i. Either, there is no world in any sphere in $\$_i$ at which ϕ, i.e., no ϕ-world, or there is no sphere $\mathcal{S}$ in $\$_i$ with at least one ϕ-world such that $\phi \supset \psi$ is true at every world in $\mathcal{S}$. (1) entails all counterfactuals with impossible antecedents are vacuously true, e.g.,

(3) If King Charles II had been my father, I would have been a royal.

1.3. Note that this account gets the failure of monotonicity, transitivity, and contraposition right.

Failure of Monotonicity: Monotonicity fails, since if $\phi \Box\!\!\rightarrow \psi$ is true at i because some $\mathcal{S}$ including a ϕ-world is such that $\phi \supset \psi$ holds at every world in $\mathcal{S}$ doesn't automatically entail that there is some sphere $\mathcal{S}'$ including a $(\phi \wedge \chi)$-world is such that $(\phi \wedge \chi) \supset \psi$ holds at every world in $\mathcal{S}'$.

Failure of Contraposition: Suppose $\phi \Box\!\!\rightarrow \psi$ and ϕ is entertainable. Thus, there is some sphere $\mathcal{S}$ such that there is at least one ϕ-world and $\phi \supset \psi$ is true at all worlds in $\mathcal{S}$. That doesn't entail that there is some sphere $\mathcal{S}'$ with a $\neg\psi$-world and $\neg\psi \supset \neg\phi$ is true at all worlds in $\mathcal{S}'$.

Failure of Transitivty: Suppose $\phi \Box\!\!\rightarrow \psi$ and $\psi \Box\!\!\rightarrow \chi$ and suppose both ϕ and ψ are entertainable. It follows then that there are two sphere $\mathcal{S}$ and $\mathcal{S}'$ and $\mathcal{S}$ both contains one ϕ-world and $\phi \supset \psi$ is true at all worlds in $\mathcal{S}$ and $\mathcal{S}'$ both contains one ψ-world and $\psi \supset \chi$ is true at all worlds in $\mathcal{S}'$. But it doesn't follow from this that there is a $\mathcal{S}''$ containing a ϕ-world and $\phi \supset \chi$ is true at all worlds in $\mathcal{S}''$.

These may be difficult to see at first. Lewis has nice diagrams to illustrate these failures in (Lewis, 1973).

3. *Spheres: The Limit Assumption and Counterfactual Excluded Middle*

3.1. Why all this talk of *spheres* of similar worlds around a given world? It might be thought simpler to hold that $\phi \mathrel{\Box\!\!\rightarrow} \psi$ is true just in case the closest *single* ϕ-world is a ψ-world. There are two problems with this. First, any two worlds may be *tied* for being most similar. There thus is no closest single world. This has a bearing on the validity of the counterfactual excluded middle.

Counterfactual Excluded Middle (CEM): For any ϕ, ψ: $(\phi \mathrel{\Box\!\!\rightarrow} \psi) \vee (\phi \mathrel{\Box\!\!\rightarrow} \neg\psi)$

We can immediately see why (CEM) *is* valid on a semantics formulated in terms of the single closest world. However, in a sphere semantics, (CEM) fails. From $\neg(\phi \mathrel{\Box\!\!\rightarrow} \psi)$, it only follows that there is a sphere $\mathcal{S}$ containing at least a ϕ-world but not all worlds in $\mathcal{S}$ are such that $\phi \supset \psi$. It is consistent with this that $\phi \supset \neg\psi$ is not true at all worlds in $\mathcal{S}$. (CEM) is seen to explain why the negation of a counterfactual $\phi \mathrel{\Box\!\!\rightarrow} \psi$ is often taken to be equivalent to $\phi \mathrel{\Box\!\!\rightarrow} \neg\psi$. Thus, we naturally say:

> BOB: If I had been there, the vase would not have been stolen.
>
> BILL: I disagree: If you had been there, the vase would have been stolen!

However, Lewis sees the failure of (CEM) as a virtue. If (CEM) holds we cannot truly say:

> It is not the case that if Bizet and Verdi were compatriots, Bizet would be Italian; and it is not the case that if Bizet and Verdi were compatriots, Bizet would not be Italian [he'd be French]; nevertheless, if Bizet and Verdi were compatriots, Bizet either would or would not be Italian.

3.2. A second difficulty is that there may not be *the closest* world which is the most similar world. Lewis's truth conditions are equivalent to ones given in terms of the closest sphere of worlds if only if we assume what Lewis calls the *Limit Assumption*. This is the assumption that 'as we take smaller and smaller antecedent-permitting spheres, containing antecedent-worlds closer and closer to *i*, we eventually reach a limit: the *smallest* antecedent-permitting sphere, and in it the *closest* antecedent worlds.

3.2. There are reasons for rejecting the Limit Assumption. Consider a counterfactual like.

(4) If I had been taller, I would have been able to reach it.

Consider the antecedent. There are increasingly smaller spheres of more similar worlds in which I am taller than I actually am. Yet, there is no sphere containing worlds which are *most* similar to the actual world. Suppose all worlds are such that I am $(h + x)$ metres, where h is my actual height and x is a very small number. Then there is a closer world where I am $(h + \frac{1}{2}x)$ metres! We can repeat this indefinitely.

4. *Overall Similarity and Criteria for Closeness.*

4.1. Lewis's truth conditions remain a 'skeleton' unless we flesh out an account of similarity amongst possible worlds. Note that counterfactuals are highly context dependent, and so the appropriate notion of similarity will depend on context. For instance, both of the following counterfactuals are perfectly natural.

(5) If Caesar were in Korea, he would have used the atom bomb.

(6) If Caesar were in Korea, he would have used catapults.

4.2. In Lewis's original (1973) account of counterfactuals, he appeals only to overall similarity. This is problematic, given what has been called The Future Similarity Objection. To see this, consider:

(7) If Nixon had pressed the launch button, there would have been wide-scale nuclear war.

(7) seems perfectly natural. However, as Kit Fine notes, on Lewis's account, (7) is '...very likely false. For given any world in which antecedent and consequent are both true it will be easy to imagine a closer world in which the antecedent is true and the consequent false.' (Fine, 1975). For instance, a world where Nixon presses the button *but* the button malfunctions, or the rockets malfunction, etc.

4.3. In response, Lewis first appeals to context sensitivity. That is, we should distinguish between (7) in a context where we hold nothing else fixed—we merely suppose that Nixon presses the launch button—and a context where we suppose 'our worst fantasies about the button', i.e., it works. In the first context, (7) is false, in the second context, (7) is true. However, Lewis recognises that we must have some more systematic way of judging closeness of worlds. He proposes several criteria for weighting differences between worlds.

(i) It is of the first importance to avoid big, widepsread violations of law.

(ii) It is of the second importance to maximize the spatiotemporal region which matches particular fact.

(iii) It is of the third importance to avoid even small, localized, simple violations of law.

(iv) It is of little or no importance to secure approximate similarity of particular fact.

4.4. The idea: when comparing two worlds w and v to the actual world, consider first (i). If w or v alone involves widepsread violation of law, then the other is closer. If neither, move down to (ii). If w or v alone involves a smaller region matching particular fact, then the other is closer. If neither, move down to (iii). And so on. We can see these in action by considering (7) again and the following worlds. w_0 is the actual world where Nixon does not press the button. The following worlds diverge from this. Each world is deterministic.

w_1: Just before the button is pressed, the laws of w_0 are violated in a 'simple, localised, inconspicuous way'. Nixon presses the button, bombs are launched, and there is a wide-scale nuclear war as a result.

w_2: No law of w_0 is violated, but the button is pressed and nuclear war ensues. Since no law of w_0 is violate, there must be large differences in the past, allowing for the worlds to diverge over the button pressing.

w_3: Just before the button is pressed, the laws of w_0 are violated locally. Nixon presses the button, but no bombs are dropped and no war starts because after after the button is pressed there is another small violation of law, and the launch system fails. w_3 and w_0 differ after the button is pressed.

w_4: Just before the button is pressed, the laws of w_0 are violated locally. Nixon presses the button, but then there is a wide-scale, diverse and complicated violation of the laws of w_0 which not only prevent the bombs from launching but reverse the cascading effects of the button being pressed.

(i)–(iv) explain why w_1 is closer to the actual world (w_0) than any of w_2, w_3, or w_4. First, w_2 cannot be closer to w_0 than w_1, since then we violate (ii)—at w_1 a larger spatio-temporal region matches particular fact at w_0. Second, w_3 cannot be closer to w_0 than w_1, since with w_3 we have double the violations of law, so violate (iii) *and* a smaller region matching particular facts, violating (ii). Finally, w_4 cannot be closer to w_0 than w_1, because with w_4 we have both a small violation *and* a large and complicated violation of law, contrary to (i).

References

Fine, Kit (1975). Critical Notice of Lewis, Counterfactuals. *Mind* 84, 451–458.

Lewis, David K. (1973). *Counterfactuals*. Malden, Mass.: Blackwell.