Conditionals, Part II: Philosophical Logic.

Lecture VI, *Counterfactuals and Chance*, 14th November

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

How do counterfactuals, as we've been understanding them, and chance interact?

1. *Counterfactuals and Coin Tossing*

1.1. Take Lewis's (1973) account as the paradigmatic case of an account of counterfactuals. So far, we have discussed indeterminism very little. Yet, we seem to be able to make many true counterfactual judgements about facts resulting from indeterministic processes. How should we understand these according to Lewis?

1.2. Consider the following example. Suppose a coin toss decides whether you receive a million dollars or nothing: you chose heads to win the million. The coin is tossed and it comes up tails. You lose. You say:

(1) If I had chosen tails, I would have been a millionaire (Edgington, 1995, 2003)

(1) seems like a natural claim to make. It did in fact come up tails and you would have won had that been chosen. However, landing tails is the result of an indeterministic process.

1.3. It looks like (1) is false on Lewis's view. Is this right? There are three ways we might get around this.

Response 1: You could say that (1) is indeed literally false: your choosing tails has no affect on how the coin will land and how the coin will land is what decides whether you're a millionaire. Despite this, (1) is a *reasonable assertion* when asserted in the context of (1.2): there we presuppose that the coin did land tails. The relevant context set only includes worlds in which the coin lands tails.

* Problem: this response will not be satisfactory if one is commited to the genuine truth of (1)—it seems that one could simply believe (1) and thus the grounds for *assertion* are irrelevant.

Response 2: One might think that (1) is true because the closest sphere of worlds to the actual world in which you had chosen tails is in fact the sphere where all the worlds agree that the coin lands tails. After all, it landed tails in the actual world. Thus, (1) in fact is true on Lewis's account!

* Is this in tension with Lewis's claim that 'it is of little or no importance to secure approximate similarity of particular fact'? This constraint is part of Lewis's response to the Nixon problem. That the coin actually landed tails is such a particular fact. But particular facts should still play a role.

Response 3: We should, according to this response, distinguish (1) from (1^*):

(1^*) If I had chosen tails and the coin had landed tails, I would have been a millionaire.

Whatever qualms one may have had about (1), (1^*) is certainly true. In fact, it is plausible to assent to (1^*), but not (1). When we accept (1), we really have in mind (1). Note: counterfactuals like (1) are context dependent and our judgements about them often track different ways of making them precise.

2. *Would*-Counterfactuals and Chance

2.1. There are broader worries about how counterfactuals, as we've understood them, and chance interact. Suppose a coin in which it is indeterministic how it lands and consider:

(2) If I were to toss this coin, it would come up heads.

If the coin is genuinely indeterministic, we of course shouldn't accept (2). Indeed, there seems to be something *very* wrong about thinking that (2) is true. Stalnaker notes this by contrasting the following two stories.

Story 1: Tweedledee and Tweedledum tossed a fair coin, but before they could see how it landed someone picked it up and ran away with it. Tweedledee is convinced that it landed heads, Tweedledum that it landed tails. Niether has any reason for his belief, but each still feels quite certain. Neither belief is justified, but one of them—we will never know which—is surely correct.

Story 2: This time someone ran off with the coin *before it was tossed*. Having no other coin, Tweedledee and Tweedledum argue about how it would have landed if it had been flipped. Tweedledee is convinced that it would have landed heads, Tweedledum that it would have landed tails. Again, neither has a reason—they agree that the coin was a normal one and that the toss would have been fair. This time, there is little inclination to say that one of them must be right. Unless there is a story to be told about a fact that renders one or the other of the counterfactuals true, we will say that neither is. (Stalnaker, 1984: 164–5)

Accepting (2), it would seem is making the mistake in Story 2, not Story 1. That is, if the coin is genuinely *chancy* and the probability is not tracking something subjective.

2.2. This all fits rather nicely into Lewis's framework. (2) is false on Lewis's view. Assuming coin tossing is an indeterministic process, there are will be worlds w and v in the same sphere where one coin comes up heads in w but not in v. (Note that assuming indeterminism, these worlds don't differ over violated law.) Moreover, this fits with Lewis's denial of Counterfactual Excluded Middle (CEM). It is also false that 'If I were to toss this coin, it would come up tails' and thus assuming that heads and tails are the only options, we get a failure of Counterfactual Excluded Middle: for some p, q: $\neg(p \mathrel{\Box\!\!\rightarrow} q)$ and $\neg(p \mathrel{\Box\!\!\rightarrow} \neg q)$.

2.3. Note that any counterfactual like (2) is false, regardless of how many times the coin is tossed. The essential point is that if the process is indeterministic, then there simply will not be the relevant facts for the the would-counterfactual to come out true. So, if (2) is true for this reason so is:

(3) If I were to toss this coin a billion times, at least one toss would come up heads.

But (3), in contrast to (2), might seems *quite reasonable*. Is there a way of distinguishing (3) from (2)?

2.4. Forget about Lewis for now. Here's a compelling idea. We could think that a counterfactual $p \mathrel{\Box\!\!\rightarrow} q$ is true just in case the chance of q would pass some high threshold, were it the case that p. Thus (3), but not (2) is true. Alan Hajek (2014) calls this:

If High Chance Then Would: $p \mathrel{\Box\!\!\rightarrow} q$ is true just in case $p \mathrel{\Box\!\!\rightarrow} C(q) > t$, where $C(q)$ is the chance of q and t is some sufficiently high threshold value. (For example, 0.999999.)

2.5. This won't work. Suppose we have a coin with a very high chance of landing heads—high enough to exceed the important threshold t. For the sake of argument, suppose that the chance of landing heads is 0.9999999. It follows from this that (4) holds. From the above, this entails (5):

(4) If I were to toss this (weighted) coin a billion times, C(the **first toss** results in heads) $> t$.

(5) If I were to toss this (weighted) coin a billion times, the **first toss** would result in heads.

Similarly, given that just the general chance that the coin landing heads is greater than t, we should accept:

(6) If I were to toss this (weighted) coin a billion times, C(the **second toss** results in heads) $> t$.

(7) If I were to toss this (weighted) coin a billion times, C(the **third toss** results in heads) $> t$.

And so on, all the way up to the billionth toss. Thus, from **If High Change Then Would**, we should accept:

(8) If I were to toss this (weighted) coin a billion times, the **second toss** would result in heads.

(9) If I were to toss this (weighted) coin a billion times, the **third toss** would result in heads.

And so on, all the way up to the billionth toss. But the chance that it lands heads every single of the billion times is $0.999999^{1000000000}$, i.e., a very small chance. Indeed, the chance that the coin *doesn't* land heads every single time is $(1 - 0.999999^{1000000000})$. This chance is > 0.999999, i.e., greater than our threshold t. So, from **If High Chance Then Would**, we should also accept:

(10) If I were to toss this (weighted) coin a billion times, it would not land heads every single time.

An uncontroversial principle governing counterfactuals is so-called Agglomeration. That is,

Agglomeration: If $p \mathrel{\Box\!\!\rightarrow} q_1$, $p \mathrel{\Box\!\!\rightarrow} q_2$, ..., $p \mathrel{\Box\!\!\rightarrow} q_n$, then $p \mathrel{\Box\!\!\rightarrow} (q_1 \wedge ... \wedge q_n)$.

By **Agglomeration**, the reasoning above entails that we should accept:

(X) If I were to toss this (weighted) coin a billion times, the **first toss** would results in heads, the **second toss** would result in heads, the **third toss** would result in heads, ..., the **billionth toss** would result in heads and it would not land heads every single time.

This we cannot be accept. Thus, we either reject Agglomeration (and Lewis's and Stalnaker's theory) or we reject **If High Chance Then Would**. But what, then, will distinguish between (2) from (3)?

References

Edgington, Dorothy (1995). On Conditionals. *Mind* 104, 235–329. DOI: 10.1093/mind/104.414.235.

Edgington, Dorothy (2003). Counterfactuals and the Benefit of Hindsight. In: *Cause and Chance: Causation in an Indeterministic World*. Ed. by Phil Dowe and Paul Noordhof. Routledge.

Hajek, Alan (2014). *Most Counterfactuals Are False*.

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

Stalnaker, Robert C. (1984). *Inquiry*. Cambridge University Press.