305 Lecture 14.1 - Counterfactuals and Similarity

Brian Weatherson

To discuss the notion of similarity at the heart of Lewis's theory.

This isn't covered so much in the book.

Julius Caeser in Command

Which of these (if either) sounds true?

1. If Julius Caeser had commanded the US forces in Vietnam, he would have used nuclear weapons.
2. If Julius Caeser had commanded the US forces in Vietnam, he would have used catapults.

Lewis argued that either could be true, in different contexts, though they couldn't both be true at once.

- A world where Caeser is in command in Vietnam is going to be very different to reality in several ways.
- Making it similar requires highlighting some aspects of similarity, and, by necessity, downplaying others.

Respects of Similarity

Lewis argued that either could be true, in different contexts, though they couldn't both be true at once.

- A world where Caeser is in command in Vietnam is going to be very different to reality in several ways.
- Making it similar requires highlighting some aspects of similarity, and, by necessity, downplaying others.
- Hold fixed how belligerent Caeser is, and you get him using nuclear weapons.

Respects of Similarity

Lewis argued that either could be true, in different contexts, though they couldn't both be true at once.

- A world where Caeser is in command in Vietnam is going to be very different to reality in several ways.
- Making it similar requires highlighting some aspects of similarity, and, by necessity, downplaying others.
- Hold fixed how belligerent Caeser is, and you get him using nuclear weapons.
- Hold fixed his knowledge of weaponry, and you get him using catapults.

Conjoining

Could this be true?

- If Julius Caeser had commanded the US forces in Vietnam, he would have used catapults to fire nuclear weapons.

Feature, Not Bug

It feels like in some sense the world in which Caeser uses nuclear weapons is more similar to actuality, and in some other sense the world in which he uses catapults is more similar. Lewis argued that this was a good feature of his theory.

- These conditionals just are vague, and context-sensitive.
- It's a good thing that we use a vague, content-sensitive notion like **similarity** to analyse them.
- That doesn't make them too vague, or context-sensitive, it **explains** why they are so vague and context-sensitive.

Feature, Not Bug

It feels like in some sense the world in which Caeser uses nuclear weapons is more similar to actuality, and in some other sense the world in which he uses catapults is more similar. Lewis argued that this was a good feature of his theory.

- These conditionals just are vague, and context-sensitive.
- It's a good thing that we use a vague, content-sensitive notion like **similarity** to analyse them.
- That doesn't make them too vague, or context-sensitive, it **explains** why they are so vague and context-sensitive.

I'm not sure whether that argument works, but I'm not going to investigate it further.

A Different Example

This seems like a tricky to figure out, historical question.

1. If Richard Nixon had ordered the use of nuclear weapons in Vietnam, his generals would have refused the order.

A Different Example

This seems like a tricky to figure out, historical question.

1. If Richard Nixon had ordered the use of nuclear weapons in Vietnam, his generals would have refused the order.

You can imagine trying to figure that out by careful archival research, and guesswork about the parts of the archives that are off-limits to researchers.

A Different Example

This seems like a tricky to figure out, historical question.

1. If Richard Nixon had ordered the use of nuclear weapons in Vietnam, his generals would have refused the order.

You can imagine trying to figure that out by careful archival research, and guesswork about the parts of the archives that are off-limits to researchers.

- But there's an argument that philosophy shows 1 is true.

Two Worlds

- Let w be the actual world where (I'll assume), Nixon never gave such an order.
- Let w_1 be the world where the order is given, and refused, and we never hear about it, and life goes on more or less as in w.
- Let w_2 be the world where the order is given, carried out, and other relevant parties (especially Russia and China) react in ways you'd expect to nuclear weapons being launched against their ally/neighbor.

Which of these worlds is more similar to the actual world?

Two Worlds

Which of these worlds is more similar to the actual world?

- There is a good case for w_1.
- It just requires a few generals to act a little out of character for a few minutes.
- In w_2, everyday life is changed in unrecognisable ways for millions of people; perhaps billions of people if it triggers a larger nuclear war.
- So philosophy tells us that if Nixon had ordered nuclear weapons to be used, his generals would have refused.

This is an absurd result.

- Maybe Nixon's generals would have refused.
- But you can only tell that by careful historical research, not by thinking about similarity of worlds.
- Something must have gone wrong.

Fixing The Problem

Insist that similarity of patterns or regularities is more important than similarity of particular facts.

- In w_1, some generals (arguably) do something very out of character.
- That's the kind of violation of a pattern or regularity that makes a big difference in similarity. (In the special sense of similarity that we care about.)
- In w_2 there are lots of particular facts that are different. There is a lot more radiation poisoning in Vietnam, and possibly in the United States (depending on the retaliation), but the patterns or regularities are not violated.

Which of these is true?

1. If I'd jumped out my office window, I would have been seriously injured (I'm 1.5 floors above some concrete steps).
2. If I'd jumped out my office window, I would only have done that if there were a net to catch me, so I wouldn't have been seriously injured.

- At least some of the time, we want to say that 1 is true.
- That's the 'non-backtracking' reading of the conditional that's relevant for thinking about decision-making, moral responsibility, causation, etc.
- But it's not clear Lewis can get that result.

Three Worlds

- World w is the actual world where I stay nice and safe in my office working on these slides.
- World w_3 is where there is no net, I jump and am seriously injured.
- World w_4 is where I check that there is a net, and then jump.
- World w_4 differs from w in two respects – there is a net that doesn't really exist, and I jump out the window.
- But world w_3 differs in two ways as well – I jump out the window, and this is **really** out of character

The New Puzzle

The 'solution' to the previous case seems to make this worse.

- If we prioritise patterns and regularities, then it looks like w_3 is really dissimilar to the actual world.
- But we want there to be an ordinary sense in which it's just true that if I had jumped out the window, I would have been seriously injured.

- If you use an overall, intuitive, measure of similarity, you get the wrong results in the nuclear weapons example.
- If you use a measure that focuses on patterns and regularities, you get the wrong result (or at least rule out a good result) in the jump out window example.
- This is starting to feel like a trap.

Here's how Lewis suggested getting out of the trap. (I'm simplifying a bit here; if you're interested you should read his paper "Counterfactual Dependence and Time's Arrow".)

- Treat the past and the future differently.
- In terms of the past, what matters for similarity is keeping **everything**, patterns, regularities, particular facts, everything, **exactly the same**. The nearest worlds are an exact duplicate of ours from the Big Bang to as close as possible to the present.
- In terms of the future, patterns and regularities are **much** more important.

Intuitive Picture

How to find the nearest world where A is true.

- Start at the present time and work backwards.
- What's the first time where you can make A true with only a single violation of patterns or regularities in the world. It might be a big violation – like gravity stopping for a second, but ideally it is just one violation.
- Roll the world forward from that time according to the laws of the world – the physical laws, the biological laws, the psychological laws, the economic laws and so on.
- If those laws leave lots of things open, then there are lots of worlds that are equally close to actuality where A is true.

Two Puzzles

I like this intuitive picture a lot – I think it captures a lot of what we're trying to do with counterfactuals. But there are still puzzles remaining. I'll end this lecture with two of them.

1. The car theft puzzle.
2. The baseball jinx puzzle.

- Imagine that yesterday I parked my car at work in a UM garage, then drove home at the end of the day.
- My car wasn't stolen yesterday (like every other day).
- But car thefts happen – it being stolen isn't like aliens invading.

The Car Theft

- Imagine that yesterday I parked my car at work in a UM garage, then drove home at the end of the day.
- My car wasn't stolen yesterday (like every other day).
- But car thefts happen – it being stolen isn't like aliens invading.
- Think about this counterfactual:

  If my car had been stolen yesterday, it would have been stolen just before midnight.

- Intuitively, that's not true.
- Just before midnight, my car was locked in my garage.
- During the day, it was in a public carpark at UM.
- If it was going to get stolen yesterday, it's at least as likely that it would have been stolen during the day as late at night.

Three Worlds (Again)

- Again, let w be the actual, no car theft, world.
- Let w_5 be the world where my car is stolen from a UM parking lot at midday.
- Let w_6 be the world where my car is stolen from my garage just before midnight.
- In w_6, things stay exactly the same as in w for much longer – for nearly 12 hours longer.
- So by the metric we're using, it's more similar to actuality than w_5.
- So it will be true that had my car been stolen yesterday, it would have been stolen just before midnight.
- This isn't very plausible, so we still need to tinker with the similarity metric.

The Baseball Jinx

Imagine that I watch a baseball game, and my team loses. My friends accuse me of jinxing the team. I defend myself with 1.

1. If I hadn't watched, they would still have lost.

Imagine that I watch a baseball game, and my team loses. My friends accuse me of jinxing the team. I defend myself with 1.

1. If I hadn't watched, they would still have lost.

Here's an argument that 1 is, for all we know, false.

Physics and Baseball

- Possibly baseball is seriously indeterministic.
- Possibly quantum indeterminacy in the player's brains causes differences in outcomes.
- Possibly quantum indeterminacy in the interaction of the ball with the air as it travels to the plate causes just enough variation in the position of the ball when it's hit to make the difference between a home run and a fly ball at the wall.
- Maybe physics will tell us otherwise, but it seems to me we can't rule out this kind of indeterminacy.

Now compare these two worlds (to w, the actual world).

- In w_7, I don't watch, but the game goes **exactly** as it goes in w, and my team loses.
- In w_8, things go **exactly** as they do in w up until the start of the game, when I don't sit down to watch it. Then they follow the laws of nature. But the laws are chancy, and in w_8 the variations favor my team just enough that they narrowly win rather than narrowly lose.

Now compare these two worlds (to w, the actual world).

- In w_7, I don't watch, but the game goes **exactly** as it goes in w, and my team loses.
- In w_8, things go **exactly** as they do in w up until the start of the game, when I don't sit down to watch it. Then they follow the laws of nature. But the laws are chancy, and in w_8 the variations favor my team just enough that they narrowly win rather than narrowly lose.

- If all we care about is exact match before the 'break point' (when the first thing changes), then conformity to laws, patterns and regularities after the 'break point', then w_7 and w_8 are equally close to actuality.
- They both are exactly alike until I don't watch (rather than watch) and both conform to laws afterwards.
- So the conditional, If I hadn't have watched, we still would have lost, is false.
- In one of the nearest worlds where I don't watch, we lose.

This doesn't sound great.

- We know that we don't make a difference to baseball games by watching or not watching them.
- But this theory, which otherwise looks promising, seems to say that maybe we do.
- This is just an open puzzle.

We'll talk more about the logic of these counterfactuals.