II. More on Subjunctive Conditionals

Subjunctive Conditional Fallacies

To get a grip on how subjunctive conditionals work it's good to think about the logic they generate. There are a number of valid inference patterns associated with the material conditional which are not valid for the subjunctive conditional, given the Lewis/Stalnaker semantics. We shall examine the three most discussed.

1. Strengthening the Antecedent

The material conditional permits strengthening of the antecedent, in the sense that all arguments of the form

(P → Q)

Therefore ((P ∧ R) → Q)

are valid.

The same is not true of subjunctive conditionals. Consider the argument

> If the Democrats had not won the last presidential election, the Republicans would have won.
>
> Therefore, if the Democrats had not won the last presidential election and the Communist Party had got ninety per cent of the popular vote, the Republicans would have won.

That is clearly not a good argument. If the Communist Party had got ninety per cent of the popular vote, then, despite the oddities of the electoral college they would have won the election. The Lewis/Stalnaker account of subjunctive conditionals explains this fact. The truth of (P □→ Q) requires that the nearest P-world be a Q-world; but the nearest P-world might not be an R-world. To find the nearest world that is P and R we might have to move to a still more distant world. And that world might not be a Q-world.

However, there are some cases where strengthening the antecedent looks like a good move. Consider the inference pattern know as 'simplification':

((P ∨ R) □→ Q)

Therefore (P □→ Q)

That is a kind of strengthening of the antecedent, and seems right for subjunctive conditionals (if I were to go to France or to Italy my phone would work, so if I were to go to Italy my phone would work, etc.), but the Lewis-Stalnaker account rejects it.

For an account that endorses it, phrased not in terms of possible worlds but in terms of possible *states*, see Kit Fine, 'Counterfactuals Without Possible Worlds', *Journal of Philosophy* 109, (2012). The idea, roughly, is that for ((P ∨ R) □→ Q) to be true, the states that verify the antecedent must exactly verify the consequent; each of them must do so on its own.

(Note, accepting Simplification is not obviously right: (1) If Spain had fought with the Axis or the Allies, she would have fought with the Axis does not appear to imply (2) If Spain had fought with the Allies, she would have fought with the Axis. Fine has to reject this using a principle that says we try to interpret counterfactuals in a way that makes their antecedents genuine possibilities.)

2. Transitivity

The material conditional is transitive, in the sense that the following inference pattern is valid:

(P → Q)

(Q → R)

Therefore (P → R)

In contrast subjunctive conditionals are not transitive. Consider the argument

If J Edgar Hoover had been born a Russian, then he would have been a communist.

If J Edgar Hoover had been a communist, then he would have been a traitor.

Therefore, if J Edgar Hoover had been born a Russian, then he would have been a traitor.

Again that's not a good argument: if Hoover had been born a Russian he would have been a patriotic communist. Again the Lewis account of subjunctive conditionals explains why not. (P □→ Q) requires that the nearest P-world be a Q-world; and (Q □→ R) requires that the nearest Q-world be an R-world. But it is consistent with both of those facts that the nearest Q-world is closer than the nearest P-world. And if that is so, the nearest P-world might fail to be an R-world.

3. Contraposition

As a final example of a subjunctive conditional fallacy, consider the inference pattern:

(P → Q)

Therefore (¬ Q → ¬ P)

This is valid. But once again the same does not hold for subjunctive conditionals. Consider:

If Boris had moved into the house, then Olga would not have moved out.

Therefore, if Olga had moved out of the house, then Boris would not have moved in.

That argument is not valid. We can easily describe a state of affairs that makes the conclusion true and the conclusion false. Suppose that Olga wanted to live in the same house as Boris, but the sentiment was not reciprocated. Had Boris moved into the house in which Olga was living, Olga would have been delighted and would have stayed on. (The premise is true.) However, the house itself was a very nice one: Boris wanted to move into it, and was only put off doing so by Olga's presence. (The conclusion is false.)

The Lewis-Stalnaker account explains why subjunctive conditionals don't contrapose, that is, why (P □→ Q) doesn't entail (¬Q □→ ¬P). (P □→ Q) requires that the nearest P-world be

a Q-world. If the nearest Q-world were nearer than the nearest P-world, then it would follow that (¬Q □→ ¬P). But it could be that the nearest ¬Q-world is further away still (i.e. further away than the nearest P-world). But then it would not follow that such a world must be a ¬P-world, and so it wouldn't follow that (¬Q □→ ¬P).

Different similarity measures

We mentioned above the fact that similarity measures are vague. Some examples that seem to show that different subjunctive conditionals sentences might require different conceptions of which worlds are more similar to the actual world. Thus consider the sentences:

If Boston were in Florida, then Boston would be in the South.

If Florida included Boston, then part of Florida would be in the North.

Both of these sentences seem to be true. If so, it seems that in assessing the first we imagine a world in which we keep Florida's borders where they are, and move Boston within them; this is the closest world in which the antecedent is true. In assessing the second we imagine leaving Boston where it is, and moving Florida's borders to include it; this is closest world in which the antecedent is true. Yet it might look as though the two antecedents say the same thing. Somehow the words used indicate that they don't; different similarity measures are required. Clearly it will be no easy thing to say exactly why this happens.

Time's Arrow and Backtracking

There seems to be a problem with this account. Consider the counterfactual

(N) If Nixon had pressed the nuclear button in 1972 there would have been a nuclear war

That looks to be true. But the world in which there would have been a nuclear war in 1972 is very different to the actual world. If we go to the world that is most similar to the actual world, but in which Nixon pushed the button, it is a world in which others then overrode him, or some-such. Then there would have been no nuclear war, and things could have continued as they actually did. But then according to Lewis's account, (N) would seem to come out false, when intuitively we thought it was true. (If you don't like the example, substitute another with a similar structure that you take to be true that would have had significant consequences: "If I'd decided not to come to Cambridge I'd have …" etc.)

Lewis's response is to say that 'closeness' and 'similarity' are technical terms here, and we should understand them by looking at how we intuitively evaluate counterfactuals. The rules of thumb that he arrives at in evaluating closeness in this way are:

Most of all, avoid big, widespread, violations of the laws of nature (big miracles);
Then try to maximize the spatiotemporal regions that have no differences whatsoever;
Then try to avoid as many small violations of the laws of nature as possible (small miracles);
But don't try to maintain approximate similarity in the spatiotemporal regions that are different.

If determinism it true, then, if the antecedent of a counterfactual is in fact false, then any world which has the same past, but in which the antecedent is true, will have to be one in

which a small miracle occurs (a miracle from our point of view: from the point of view of that world, the laws will be different). In fact even if determinism isn't true, that will still typically be the case. So in evaluating counterfactuals, we will have to be constantly conceding small miracles.

Given the way that the laws of nature actually work, Lewis thinks that this account gives rise to a temporal asymmetry. When an event happens, it has many consequences. So if we try to undo that event, i.e. put things back so that it was as though it never happened, we will have to propose many miracles. (Imagine the ripples radiating out from a stone dropped in a pond; the longer things go on, the more has to be undone to get things back as they would have been had it not been dropped.) That, Lewis thinks, explains the Nixon case: it would take lots of miracles to put things back again, once Nixon had pressed the button. Given his similarity measure, the closest world in which Nixon presses the button is the one in which the only miracle is the one that was needed for him to press it.

This explains another interesting feature of counterfactuals: they don't normally allow backtracking. That is, when we consider the truth of a counterfactual, we standardly try to keep things prior to the event mentioned in the antecedent as close to the actual world as possible. In contrast things after the event mentioned in the antecedent will unfold however the laws in that world will lead them to unfold—and the laws will be kept as close to the laws in the actual world as possible. Suppose you're waiting in the queue for lunch. I say: 'If you miss lunch you'll be starving by mid-afternoon'. It may be that if you were to have missed lunch, the most likely explanation would have been that you were off your food for some reason; and if you were off your food, you wouldn't be starving by mid-afternoon. But that doesn't invalidate the truth of the counterfactual. We typically take things as they actually are to the point at which you would have lunch; and then we imagine you missing it, and see how things work out. We don't want to build this into the truth conditions of the counterfactual though: we don't want to insist by fiat that things must be kept the same prior to the antecedent. Imagine how we would evaluate a story containing time travel; that would typically have to involve other prior differences. It is just that in keeping to the rules of similarity, we will not generally allow backtracking, since that will require greater divergence. But we can force a backtracking reading by context, and typically we will use a special form of words: If you were to have missed lunch, you would have had to have been off your food. Again, different languages do it differently. (For discussion of all this, see Lewis's 'Counterfactual Dependence and Time's Arrow')