I. Introduction, & Subjunctive Conditionals I

The webpage for the course:

https://rjh221.user.srcf.net/courses/conditionals/

Consider the two sentences

(1) If Oswald didn't shoot Kennedy, then someone else did. (did-did)

(2) If Oswald hadn't shot Kennedy, then someone else would have (had-would)

Clearly these don't mean the same thing. The first has some claim to be read as the material conditional. All that is clearly ruled out is the possibility that the antecedent is true (i.e. Oswald didn't shoot Kennedy) and the consequent is false (i.e. nobody else shot him either). But the second sentence cannot be read as a material conditional. The fact that the antecedent is false (since, let us suppose, Oswald did shoot Kennedy) doesn't, by itself, make the sentence true. So it looks as though there are two quite different 'If..., then' constructions in English, marked by the different mood of the verbs involved. In (1) the verbs are in the simple indicative mood; in (2) they are subjunctive ('had shot', 'would have shot').

Following fairly standard usage, we'll call sentences of the form of (1) 'indicative conditionals'; and we'll call sentences of the form of (2) 'subjunctive conditionals.' An alternative, equally, popular term for the latter is 'counterfactuals', which we will sometimes use. But that has the disadvantage that it suggests that the antecedent is counter to fact, which, for reasons we'll see shortly, it need not be. For a long time is was widely assumed that the material conditional, perhaps with a little pragmatic supplementation, gave the right account of the indicative conditional. Very few people now believe that. We'll start with subjunctive conditionals, and look at indicative conditionals—and whether they can be understood on the model of the material conditional—later.

Note before starting that this simple division into two kinds of conditional, whilst widely accepted, is not universally held. For a start, there is a question of how to classify certain sentences:

(3) If you swim in the sea today your cold will get worse (does-will, from Bennett)

Secondly, there are a few people (Vic Dudman is the most influential) who deny the whole distinction, and wish to replace it with one based on tense. We won't investigate this (see Bennett, pp.14–15 for discussion and references).

Truth Conditions for Subjunctive Conditionals

We'll symbolize subjunctive conditionals as follows:

$(P \mathbin{\Box\!\!\rightarrow} Q)$ (this is from Lewis; an alternative, from Stalnaker, is '>')

Note that this is emphatically not the same as the strict conditional:

$\Box (P \rightarrow Q)$

In developing truth conditions for counterfactuals we follow the account given independently by David Lewis and Robert Stalnaker (anticipated a few years earlier by one William Todd from Cincinnati, about whom I know nothing more) who say (roughly):

> (P □→ Q) is true (at the actual world) iff the closest possible world (i.e. closest to the actual world) in which the antecedent, P, is true, is a world in which the consequent, Q, is also true (or, in other words, (P □→ Q) is true iff the closest P-world is a Q-world).

What do we mean here by 'closest'? Lewis glosses it as a measure of similarity. The closest P-world to the actual world is the world in which P is true which is most similar to the actual world (although even this is a technical term: for reasons we'll see when we talk about backtracking it certainly doesn't just involve comparing each atomic fact in the world (whatever those might be) and looking for the best match). So the account of subjunctive conditionals amounts to this: a subjunctive conditional (P □→ Q) is true just in case the world most similar to the actual world in which P is true is a world in which Q is true. This means in order to assess the truth value of a subjunctive conditional we have to make an assessment about similarities between worlds; and that is going to be a rather vague business. But we shouldn't let that put us off the account. The truth value of subjunctive conditionals is itself vague; the account should mirror that vagueness.

No other world can be as similar to a world as that world is to itself. Identity is the limit case of similarity. But if that is so, then, if the actual world is a P-world, (P □→ Q) will be true just in case the actual world is a Q-world. That might seem to be wrong: surely we would never say 'If Oswald hadn't shot Kennedy, someone else would have' if we knew that in fact Oswald hadn't shot him. But, as ever in providing a semantics for natural language, we need to distinguish that which is false from that which is pragmatically unacceptable on other grounds. It is true that we would normally not utter a subjunctive conditional if we knew that its antecedent was true; but that could be because, in such circumstances, we would be in a position to assert the consequent itself, and so it would be misleading to assert something weaker. You wouldn't say 'If they were to find out, you'd be in big trouble' if you knew they had found out; you'd just say: 'They've found out. You're in big trouble!' This doesn't show that the subjunctive conditional would be false. Indeed there are good reasons for thinking that it would not be. Consider this exchange:

> A: If they were to find out, then you'd be in big trouble
>
> B: Damn! I've already told them!

Here B doesn't deny what A says, on the grounds that it's a counterfactual whose antecedent is true. Quite the reverse: B uses A's counterfactual to reach the conclusion that she is in trouble. So it seems reasonable to assume that the Lewis account is right: counterfactuals with true antecedents are true just in case their consequents are true. The reason that we don't typically assert them is pragmatic.

Stalnaker's Assumption and Conditional Excluded Middle

We said that this account was roughly that given by Lewis and Stalnaker; in fact it is closer to Stalnaker's. We have simplified Lewis's account in a number of ways. The most significant concerns our talk of *the* closest P-world. Stalnaker argues that, provided P is not contradictory (in which case the conditional is vacuously true), there will always be such a world. Lewis argues that, even if P is not contradictory, there are two ways in which there might fail to be such a world, and yet the counterfactual still be true. First, there might be two or more P-worlds that are equally close; provided that these worlds are all Q-worlds, that shouldn't make the counterfactual come out false. Second, there might be an infinite series of P-worlds, each one of which is closer to the actual world than the one before—compare the infinite series of fractions 1/2, 1/4, 1/8, 1/16 ... each of which is closer to zero that the one that comes before. (To accept this possibility is to deny what Lewis calls the *Limit Assumption*.) Again, provided that these are all Q-worlds, the counterfactual should still be true. Lewis avoids these problems by saying that $(P \mathbin{\Box\!\!\rightarrow} Q)$ will be true iff there is a possible world, w, which is both a P-world and a Q-world, and that any P-world which is as close or closer to the actual world than w is also a Q-world. It's not so easy to get one's mind around this formulation; so in our discussion we'll stick with our simpler approximation. But it's worth examining what's at issue.

We can think of Stalnaker's account as the special case of Lewis's account that results when we add two further assumptions: the Limit Assumption and the No-Ties Assumption (Lewis calls the conjunction of these 'Stalnaker's assumption'). Most of the debate has concerned the No-Ties Assumption: that there cannot be two or more worlds that are equally close. There are two connected advantages in making that assumption. First, it enables us to accept as a theorem Conditional Excluded Middle (CEM):

$$(P \mathbin{\Box\!\!\rightarrow} Q) \vee (P \mathbin{\Box\!\!\rightarrow} \neg Q)$$

If we deny No-ties, that won't a theorem, since one of the closest P worlds may be a Q world, while the other is a not-Q world. But a lot of people think that CEM has some intuitive force of its own, and it has some nice consequences: it makes $(P \mathbin{\Box\!\!\rightarrow} \neg Q)$ and $\neg (P \mathbin{\Box\!\!\rightarrow} Q)$ equivalent, which seems intuitively right.

Second it means that we won't get into the rather odd positions of denying both $(P \mathbin{\Box\!\!\rightarrow} Q)$ and $(P \mathbin{\Box\!\!\rightarrow} \neg Q)$, but accepting $(P \mathbin{\Box\!\!\rightarrow} (Q \vee \neg Q))$

An example, originally from Quine:

(5) If Bizet and Verdi were compatriots, Bizet would be Italian

(6) If Bizet and Verdi were compatriots, Bizet would not be Italian (he would be French)

(7) If Bizet and Verdi were compatriots, Bizet either would or would not be Italian

Lewis has to reject both (5) and (6) as false, but accept (7).

What about Stalnaker though? Does he have to say that one of (5) and (6) is true and the other false? If so, which? (The implausibility of preferring one to the other was Quine's original reason for giving the examples; he used it to reject the whole idea of counterfactuals as hopelessly vague.)

In fact Stalnaker wants to accept neither (5) nor (6) as true; but neither does he want to say that they are false. Instead he wants to say that they are indeed vague, and that they should be treated using a standard account of vagueness, namely supervaluation. The idea is that that there are two equally good ways of making precise the relevant similarity relation (two ‘valuations’), one that makes (5) true and (6) false, and one that makes (6) true and (5) false. Under those circumstances, both (5) and (6) should be treated as neither true not false. But both make (7) true (it is true on the ‘supervaluation’), so it is true simpliciter.

Centering

A second issue concerns whether we should have an account in which (P □→ Q) follows from the truth of P and Q. That follows from the assumption that no world is *as similar* to the actual world as the actual world. If we remove that assumption (but keep the assumption that no world is *more similar* to the actual world than the actual world) we arrive at the hypothesis of weak centering, and (P □→ Q) no longer follows from the truth of P and Q.

Why might we want to do this? The basic idea is that counterfactuals should be robust: if changing the actual world a little bit would result in the antecedent staying true whilst the consequent becomes false, that might not seem good enough for a the truth of the counterfactual. (Similarly, if the antecedent is false in the actual world, we might want to say that for the counterfactual to be true, it’s not enough that the consequent be true at the closest antecedent world, but that it must be true across a band of close antecedent worlds.) But when we apply this to actual case the results are often odd: consider, for instance what a counterfactual account of causation or of moral responsibility would look like that used such a conditional.

Further Reading

David Lewis, *Counterfactuals* (Blackwells, 1973).

For Stalnaker’s account see ‘A Theory of Conditionals’ and in defence of the ways in which it diverges from Lewis’s, ‘A Defense of Conditional Excluded Middle’.

Jonathan Bennett, *A Philosophical Guide to Conditionals* Chs. 10–21