CONDITIONALS LECTURE 3

1. Subjunctive conditionals take the form 'If it had been the case that p then it would have been the case that q', which I write p > q. Lewis (in his 1973 book of that title) calls these 'counterfactuals' because one who asserts them typically believes, and intends somehow to get across, that the antecedent is in fact false; there are however examples where we knowingly assert subjunctives with true antecedents (see e.g. Anderson's article in *Analysis* 1951).

2. It is worth noting three logical facts about the subjunctive that distinguish it from both the material conditional and strict implication:

<u>Failure of monotonicity</u>: from p > q we cannot in general infer (p&r) > q for arbitrary r (contrast with entailment and strict implication). Example: p= 'I jump out of a high building', q= 'I break my leg', and r= 'I am rescued by an alien spaceship on my way down'.

<u>Failure of contraposition</u>: from p > q we cannot in general infer ~q > ~p. Example: p= 'Oswald didn't kill Kennedy' and q= 'Kennedy was in Texas on 22 November 1963'. If Oswald hadn't killed Kennedy, then Kennedy would still have been in Texas on that day; but if Kennedy hadn't been in Texas on that day then Oswald would never have had the chance to kill him.

<u>Failure of transitivity</u>: from p > q and q > r we cannot infer p > r. Example: take p= "Hoover is Russian", q= "Hoover is a communist", r= "Hoover is a traitor".

Any semantic theory of counterfactuals should at least accommodate and at best explain these facts.

3. The first serious attempt at this is Goodman's paper in F. Jackson, ed., *Conditionals* (also ch. 1 of Goodman's *Fact, Fiction and Forecast*). Goodman is trying to say something about the criteria for whether a subjunctive is true or false. Consider the proposition that if this match had been struck then it would have been lit (intuitively it is true, even though we know very well that in fact the match was *not* struck). Goodman represents our judgment about this as an inference from the antecedent, together with various propositions that are believed true, in accordance with laws of nature.

4. Thus in the match case, we infer from the supposition, that the match was struck, together with various truths about the actual situation: that the match was dry and well-made, that there was oxygen in the atmosphere, that the striking surface was rough, etc. From these facts, together with the supposition that the match was struck, we infer that the match does light, and this grounds our acceptance of the counterfactual. This suggests a theory: p > q is true iff q follows from – p given additional truths.

5. Unfortunately this doesn't work, for we don't judge the counterfactual by inferring from the supposition that the match was struck together with *any* truths. After all, one of these is the truth that the match did not light. So we should have as good grounds for asserting that *if the match had been struck then e.g. it would have been wet* as we do for asserting that if the match had

been struck then it would have lit. The only difference is that in the two cases we choose a different selection of truths as auxiliary premises. In the first case we use the true premises P1={the match was dry, it was well made, there was oxygen in the atmosphere}; in the second case we use the premises P2= {the match was well made, there was oxygen in the atmosphere, the match did not light}. In both cases the relevant law of nature is something like this: ‘All dry well-made matches will light when struck in the presence of oxygen’. Both P1 and P2 contain only true premises: what then entitles us to use P1 but not P2 as auxiliary premise when reasoning from the counterfactual supposition ‘The match is struck’? In Goodman’s terminology, we say that an acceptable auxiliary premise for counterfactual reasoning from p is **co-tenable with p**; the problem is therefore to characterize what is meant by co-tenability.

6. Goodman tried a number of criteria for co-tenability, but none of them seemed to work. His despairing conclusion was this: C is co-tenable when reasoning from the counterfactual supposition p iff C is true and would not have been false if p *had* been true. This is unacceptably circular: we are being told that p > q is true iff q is derivable from p together with laws of nature plus auxiliary true premises C s.t. C is true and ~(p > ~C). So to work out the truth-value counterfactual subjunctive, we have to work out the truth of another one. And this leads to a vicious regress. This is the ‘problem of co-tenability’. (There is a good review of the history of the problem towards the end of Jonathan Bennett’s 1984 article in *Philosophical Review*.)

7. There were several important contributions to the literature in the years after Goodman wrote, perhaps most notably Mackie’s subtle and detailed chapter on counterfactuals in *The Cement of the Universe*. But Lewis’s 1973 book is the one that has generated most discussion. Lewis’ approach is part of his general theory of ‘possible worlds’. These are concrete entities spatiotemporally disconnected from one another. But NB those metaphysical issues have little bearing on the logic of his theory of conditionals.

8. According to Lewis, we could understand possible worlds as being partially ordered by their *similarity* or ‘closeness’ to the actual world. Essentially, his view was that p > q is true iff q is true at all the p-worlds (worlds at which p is true) that are most similar to the actual world. Even without going into the details of what is meant by ‘similarity’ – except that it is some kind of ordering – we already can see how this explains the failures of inference noted above, in much the same way as Kripke’s semantics ‘explains’ such failures of modal inference as that from ‘possibly-p and possibly-q’ to ‘possibly-(p and q)’.

9. For instance, it explains the failure of monotonicity as follows: the closest p&r-worlds are further from the actual world than the closest p-worlds. (So the closest p-worlds are all p&~r-worlds). So although all of the closest p-worlds are q-worlds, and therefore p > q is true, the closest p&r-worlds are all ~q-worlds and therefore (p&r) > q is false.