

Lecture 8: Counterfactual Conditionals

1. Introduction

2. First Steps: Material and Strict Conditionals

3. Counterfactuals as Variably Strict Conditionals

...3.1. Stalnaker

...3.2. The CEM

...3.3. Lewis

Lecture 8: Counterfactual Conditionals

1. Introduction

▪ Consider the following sentences:

(1) 'If the French lose, I'll eat my beret.'

(2) 'If you hadn't been so busy arguing, then you wouldn't have missed the turning.'

▪ These sentences are known as 'conditional' sentences (or 'conditionals' for short).

▪ There are many ways of forming a conditional sentence. The canonical form is 'if p then q', where p and q are sentential phrases. Here, p is known as the 'antecedent', q is known as the 'consequent' and p is said to 'imply' q.

▪ Conditional sentences come in two main varieties.

▪ First of all we have 'counterfactual' conditionals - or simply 'counterfactuals' - in which p is said to 'counterfactually' imply q (in logical notation: $p \Box \rightarrow q$). Example:

(3) 'If Shakespeare hadn't written *Hamlet*, then someone else would have.'

Lecture 8: Counterfactual Conditionals

1. Introduction

- These conditionals are known as 'counterfactual' conditionals because the antecedent expresses a condition that is believed by the speaker to be 'contrary to fact' - i.e. false.
- Note that they are however also sometimes known as 'subjunctive' conditionals. This label is inaccurate insofar as 'subjunctive' conditionals, whether in English or in other languages, aren't in fact expressed in the subjunctive mood.
- Counterfactual conditionals are typically contrasted with 'indicative' conditionals - or simply 'indicatives' for short - in which p is said to 'indicatively' imply q . Example:
(4) 'If Shakespeare didn't write *Hamlet*, then someone else did.'
- The grammatical difference between the two varieties of conditionals reflects a semantic one: indicative and counterfactual conditional sentences which have the same antecedents and consequents can nevertheless differ in truth-values. Indeed, (3) is quite possibly false, whilst (4) is most definitely true.

Lecture 8: Counterfactual Conditionals

1. Introduction

- Today we will be talking about *counterfactual* conditionals: what makes counterfactuals true, when they are true, and what rules of inference govern correct counterfactual reasoning.
- Counterfactual conditionals?!? But isn't this a metaphysics course? Why should we care about the analysis of some obscure fragment of natural language?
- Well it turns out that counterfactual conditionals figure *heavily* in the analysis of a number of *central* metaphysical concepts. For instance, in relation to the rest of this course:
 - Laws of Nature (if it is a law of nature that all F s are G s, then had something been F , it would have been G)
 - Causation (one suggestion: if c caused e then, had it not been the case that c , it wouldn't have been the case that e)
 - Dispositions (one suggestion: x has the property of being disposed to bring about e in circumstances c iff had it been the case that c , x would have brought about e)

Lecture 8: Counterfactual Conditionals

1. Introduction

- Counterfactuals also turn up in other metaphysical topics such as free will and the philosophy of time.
- They also crop up in a large number of other fields of philosophy altogether. For instance, in epistemology, a common requirement on knowing that p is for it to be the case that had it not been the case that p , one wouldn't have believed that p .
- In what follows I will try to give you a very basic introduction to counterfactual conditionals.
- Due to time considerations, I have had to leave out some important issues. I will hopefully pick up on some of these in the lectures on causation.

Lecture 8: Counterfactual Conditionals

2. First Steps: Material and Strict Conditionals

- It is instructive to start our attempt to analyse counterfactual implication by considering a type of implication known as '*material implication*' (' p materially implies q ' is symbolised as $p \rightarrow q$, or alternatively $p \supset q$).
- Unlike counterfactual and indicative conditionals, who occur in the 'linguistic wild' and whose logic is yet to be fully understood, the material conditional is a logician's creation, whose logic was stipulatively introduced.
- By stipulative definition, $p \rightarrow q$ is true just in case, *in the actual world*:
 - (i) p is true and q is true,
 - (ii) p is false and q is false,
 - (iii) p is false and q is true.
- In other words, $p \rightarrow q$ is true iff $\sim p \vee q$ is true.

Lecture 8: Counterfactual Conditionals

2. First Steps: Material and Strict Conditionals

- It is helpful to think of the material conditional as corresponding to our natural language indicative conditional.
- Indeed indicative conditional constructions can be similarly translated into disjunctive constructions: 'if you don't do as I say, you'll be grounded' (schematically: if $\sim p$ then q) can naturally be reworded as 'either you do as I say or you're grounded' (schematically: 'either $\sim \sim p$ or q ', i.e. 'either p or q ').
- The ins and outs of the debate over the relationship between these two conditionals would take us too far afield but note that, whether or not the material conditional corresponds to the indicative conditional, it quite clearly *doesn't* correspond to the counterfactual conditional. If it did, the following sentence would come out true (because it has a false antecedent):

(5) 'Had I not come to the lecture, I would have won a million pounds.'

Lecture 8: Counterfactual Conditionals

2. First Steps: Material and Strict Conditionals

- A slightly more promising candidate is another logician's creation: the so-called 'strict conditional'. p strictly implies q iff necessarily, p materially implies q ($\Box p \rightarrow q$). (or alternatively, using possible world talk, iff, in every possible world, p materially implies q)

- This gets rid of the problem raised on the previous slide:

(5) 'Had I not come to the lecture, I would have won one million pounds.'

now comes out false: there exists, for instance, a possible world in which I don't come to the lecture but sadly don't win a million pounds.

- But here's another problem:

(6) 'Had I not come to the lecture, I would have stayed at home'

(6) seems true. However, if we take counterfactual implication to be strict implication, we have to consider (6) to be false. This is because there exist possible worlds in which I don't come to the lecture and go ice-skating instead. The intuition is here that certain possibilities are *irrelevant* in assessing the truth of a counterfactual... What to do?

Lecture 8: Counterfactual Conditionals

2. First Steps: Material and Strict Conditionals

- Here's a thought... You may remember that, in the lecture before last, I mentioned the notion of different 'strengths' of necessity. The idea was that we could obtain increasingly weaker varieties of necessity by increasingly restricting the range of possible worlds considered. Starting off with the notion of necessity simpliciter ($\Box p$), i.e. necessity in the strongest possible sense, we can obtain, for instance, the notion of nomological necessity ($\Box_{\text{Nom}} p$), by considering that proper subset of the set of all possible worlds whose members abide by the laws of nature operating in the actual world.
- One obvious response to the problem of irrelevant possibilities would be to suggest that we need to be working with a weaker notion of necessity than the notion of necessity simpliciter: a notion of necessity that rules out those problematic possibilities.
- One obvious problem with this suggestion is to find a relevant notion of necessity that will do the job. It's difficult to even think of a prima facie plausible proposal. But this, however, is comparatively minor...

Lecture 8: Counterfactual Conditionals

2. First Steps: Material and Strict Conditionals

- The real problem, according to many people, is that whatever restriction we end up settling on, it just won't work. Here's why...
- Strict conditionals (as well as, incidentally, material conditionals), whether restricted or not, can easily be proven to exhibit the following formal properties:
 - Contraposability: if $\Box p \rightarrow q$ then $\Box \sim q \rightarrow \sim p$.
 - Monotonicity: if $\Box p \rightarrow r$ then $\Box p \& q \rightarrow r$.
 - Transitivity: if $\Box p \rightarrow q$ and $\Box q \rightarrow r$ then $\Box p \rightarrow r$.
- Here's a proof for monotonicity (I'll leave the others up to you if you want to have a go):
[1] if $\Box p \rightarrow r$, then every world is either a $\sim p$ -world or an r -world. [2] every $\sim p$ -world is a $\sim(p \& q)$ -world (truth of logic). Therefore [3] if $\Box p \rightarrow r$, then every world is either a $\sim(p \& q)$ -world or an r -world, i.e. $\Box p \& q \rightarrow r$.
- Now the problem is this: intuitions seem to indicate that these formal properties aren't shared by counterfactual conditionals...

Lecture 8: Counterfactual Conditionals

2. First Steps: Material and Strict Conditionals

- The following case arguably suggests that counterfactuals violate transitivity:
 - (7) 'Had Hoover been a communist, he would have been a traitor.' (T)
 - (8) 'Had Hoover been born in the Soviet Union, he would have been a communist.' (T)
 - (9) 'Had Hoover been born in the Soviet Union, he would have been a traitor.' (F)
- There are also (in my mind rather tenuous) counterexamples to contraposition:
 - (10) 'If Goethe hadn't died in 1832, he would nevertheless be dead by now.' (T)
 - (11) 'Had Goethe still been alive now, he would have died in 1832.' (F)
- Finally, there are pretty convincing counterexamples to monotonicity:
 - (12) 'Had I stepped on the ice, it would have cracked.' (T)
 - (13) 'Had I stepped on the ice weighing less than 10 kilos, it would have cracked.' (F)
- *If* these intuitions are correct, counterfactuals cannot be strict conditionals: back to the drawing board....

Lecture 8: Counterfactual Conditionals

3. Counterfactuals as Variably Strict Conditionals

- The accepted diagnostic of the problem raised on the previous few slides is that the range of possible worlds considered when evaluating the counterfactual cannot be held fixed: it needs to vary according to the antecedent.
- In what follows, I shall examine two well-known suggestions as to how we could understand counterfactuals as *variably* strict conditionals: the accounts due, respectively, to Robert Stalnaker and to David Lewis.
- Both accounts make use of a 'closeness' metric between worlds. Of course, 'closeness' isn't intended to be read in a spatial or temporal sense. There is a fair amount of debate as to how exactly we should understand this notion. For present purposes, just think of closeness as some kind of similarity.

Lecture 8: Counterfactual Conditionals

3. Counterfactuals as Variably Strict Conditionals > 3.1. Stalnaker

- Stalnaker offers roughly the following analysis:

[S] $p \Box \rightarrow q$ iff the closest p -world to the actual world is a q -world.

- With regards to sentences (5) and (6):

(5) 'Had I not come to the lecture, I would have won a million pounds.'

(6) 'Had I not come to the lecture, I would have stayed at home'

(5) comes out as false and (6) comes out as true because the closest possible world in which I don't come to the lecture (which is, amongst other things, a world in which I don't get invited to go ice skating, a world in which I don't get abducted by aliens, a world in which I didn't play the lottery, etc.) is a world in which I stay at home and don't win a million pounds.

- This is an intuitive result. But [S] has further selling points.
- Indeed, those who buy into the various counterexamples to counterfactual monotonicity, transitivity and contraposition discussed above will be pleased to note the following: all three principles come out false according to [S].

Lecture 8: Counterfactual Conditionals

3. Counterfactuals as Variably Strict Conditionals > 3.1. Stalnaker

- With regards to transitivity (if $p \Box \rightarrow q$ and $q \Box \rightarrow r$, then $p \Box \rightarrow r$), consider:

$\sim p$
$\sim q$
$\sim r$

w_1

$\sim p$
q
r

w_2

p
q
$\sim r$

w_3

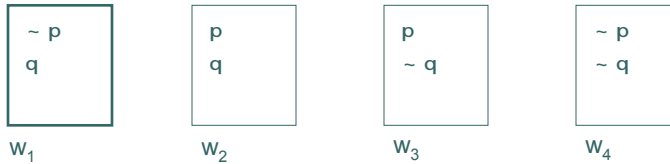
- Here we have:

- the closest p -world, w_3 , is a q -world (hence $p \Box \rightarrow q$), and
- the closest q -world, w_1 - i.e. the actual world, is an r -world (hence $q \Box \rightarrow r$), *but*
- the closest p -world, w_3 , *isn't* an r -world (hence it isn't the case that $p \Box \rightarrow r$)!

Lecture 8: Counterfactual Conditionals

3. Counterfactuals as Variably Strict Conditionals > 3.1. Stalnaker

- With regards to contraposition (if $p \Box \rightarrow q$, then $\sim q \Box \rightarrow \sim p$), consider:

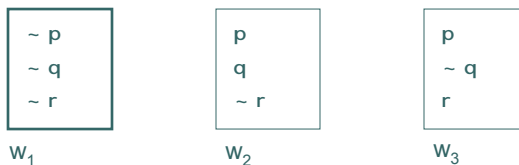


- Here we have:
 - the closest p -world, w_2 , is a q -world (hence $p \Box \rightarrow q$), *but*
 - the closest $\sim q$ -world, w_3 (and *not* w_4), is a p -world (hence $q \Box \rightarrow r$).

Lecture 8: Counterfactual Conditionals

3. Counterfactuals as Variably Strict Conditionals > 3.1. Stalnaker

- With regards to monotonicity (if $p \Box \rightarrow q$ then $p \& r \Box \rightarrow q$), consider:



- Here we have:
 - the closest p -world, w_2 , is a q -world (hence $p \Box \rightarrow q$), *but*
 - the closest $p \& r$ -world, w_3 , is *not* a q -world (hence it isn't the case that $p \& r \Box \rightarrow q$).

Lecture 8: Counterfactual Conditionals

3. Counterfactuals as Variably Strict Conditionals > 3.2. The CEM

- All this is great news (at least if you buy *all* of the intuitions about sentences (7)-(13) above), but there is controversy around the corner...
- Note that [S] assumes that for any p , there always exists a *unique* closest possible p -world (i.e. there are no ties for closeness).
- *However*, if we analyse closeness in terms of similarity, this assumption may very be false: nothing seems to prevent there from being ties for the title of 'closest' world.
- *Furthermore*, as a consequence of this uniqueness requirement, the following principle, known as the 'Conditional Excluded Middle' holds true:

CEM: $(p \Box \rightarrow q) \vee (p \Box \rightarrow \sim q)$ (i.e. 'EITHER had it been the case that p , it would have been the case that q OR had it been the case that p it wouldn't have been the case that q .')

The reason why it does is that the closest p -world must either be a q -world (making true the first disjunct) or a $\sim q$ -world (making true the second).

Lecture 8: Counterfactual Conditionals

3. Counterfactuals as Variably Strict Conditionals > 3.2. The CEM

- There is a fair amount of dispute as to whether or not this principle is true. One consideration against it is the 'argument from 'might' counterfactuals'...
- We have been talking so far about conditionals of the form 'if it were the case that p , it would be the case that q '. These are known as 'would'-counterfactuals.
- But there are also other conditionals – 'might'-conditionals – of the form 'if it were the case that p , it *might* be the case that q ' (in logical notation: $p \Diamond \rightarrow q$).
- Now it is plausible to claim that $p \Diamond \rightarrow q$ is logically equivalent to $\sim(p \Box \rightarrow \sim q)$.
- But if this is the case, CEM is in trouble. It turns out that it then trivially follows that if CEM is true, then $p \Diamond \rightarrow q$ entails $p \Box \rightarrow q$, which seems wrong. Proof:

[1] $[(p \Box \rightarrow \sim q) \vee (p \Box \rightarrow q)]$ (i.e. CEM) entails $[\sim(p \Box \rightarrow \sim q) \rightarrow (p \Box \rightarrow q)]$ (truth of logic).

[2] $(p \Diamond \rightarrow q) \leftrightarrow (\sim(p \Box \rightarrow \sim q))$ (our plausible assumption).

Therefore [3] $[(p \Box \rightarrow \sim q) \vee (p \Box \rightarrow q)]$ (i.e. CEM) entails $[(p \Diamond \rightarrow q) \rightarrow (p \Box \rightarrow q)]$.

Lecture 8: Counterfactual Conditionals

3. Counterfactuals as Variably Strict Conditionals > 3.3. Lewis

- Lewis offers an alternative that avoids commitment to CEM. Roughly:
[L] $p \Box \rightarrow q$ iff there is at least one $p \& q$ -world closer to the actual world than any $p \& \sim q$ -world.
- Why does it avoid CEM? It avoids CEM because it countenances the possibility of ties for closeness:
 - Consider a scenario in which we have two closest p -worlds: one in which it is the case that q and one in which it is the case that $\sim q$.
 - In this scenario, both $p \Box \rightarrow q$ and $p \Box \rightarrow \sim q$ turn out to be false, violating CEM.
- [L] also avoids another potential difficulty facing [S]: there may be no closest possible worlds at all, whether unique or not. It may be the case that for every non-actual world there is always a world that is closer to actuality. This poses no problem for [L].
- Finally, like [S], [L] has the consequence that counterfactuals are non-monotonic, non-transitive and non-contraposable.

Lecture 8: Counterfactual Conditionals

Next week... Laws of Nature

- Set reading: Bird, A. (1998) *Philosophy of Science*. London: UCL Press. Ch 1 'Laws of Nature'. (available on WebCT)