Handout 10: 'Might' Counterfactuals

Philosophy 691: Conditionals
Northern Illinois University ⋆ Fall 2011
Geoff Pynn

THE DUALITY THESIS

1. $\Diamond$ is usually defined as the 'dual' of $\Box$:[1]

   $\Diamond P =_{df.} \neg\Box\neg P$

   This definition is uncontroversial. Try for yourself thinking about substitution instances of 'It is possible that P' and 'It is not necessary that $\neg P$' and you'll see why.

2. Let $A \Diamond\!\!\rightarrow C \approx$ the counterfactual expressed by 'If it were the case that A, then it might be the case that C'. Lewis defines $\Diamond\!\!\rightarrow$ as the dual of $\Box\!\!\rightarrow$; we'll call this the Duality Thesis:

   DT $A \Diamond\!\!\rightarrow C =_{df.} \neg(A \Box\!\!\rightarrow \neg C)$

   Translated into the Lewis semantics, DT means that $A \Diamond\!\!\rightarrow C$ is true at α iff there is no A-sphere around α such that $A \supset \neg C$ is true at every world in that sphere. Equivalently, ignoring cases where there is no smallest A-sphere around α, $A \Diamond\!\!\rightarrow C$ is true at α iff the smallest A-sphere around α contains at least one world where $A \,\&\, C$ is true.

3. Why accept DT? Because conjunctions of the form '$A \Diamond\!\!\rightarrow C$ and $A \Box\!\!\rightarrow \neg C$' sound horrible; e.g.:

   > 'If McCain had won, the economy might be worse than it is, but if McCain had won, the economy would not have been worse than it is.'

   DT explains this datum: it says that the second conjunct is inconsistent with the first.

THE CASE AGAINST DT

1. Accepting DT requires denying CEM. (See James and Peter's handout or Bennett p. 189 for the explanation.) But denying CEM has some moderately counterintuitive consequences:

   (a) It is generally weird to deny (i) while affirming (ii).

      (i) $(A \Box\!\!\rightarrow C) \vee (A \Box\!\!\rightarrow \neg C)$

      (ii) $A \Box\!\!\rightarrow (C \vee \neg C)$

      But unless we are prepared to deny the straightforward Law of Excluded Middle (i.e., that $P \vee \neg P$ is a logical truth), we cannot deny (ii). Lewis admits that denying (i) and affirming (ii) "sounds like a contradiction," but thinks that the "cost of respecting this offhand opinion is too much" (*Counterfactuals*, p. 80). Some (not me) do not even share the offhand opinion.

   (b) The natural way to deny $A \Box\!\!\rightarrow C$ is to say $A \Box\!\!\rightarrow \neg C$. (A: "If McCain had won in 2004, we would be better off than we are." B: "No, if he had won, we wouldn't be any better off than we are.") If $A \Box\!\!\rightarrow \neg C$ is the negation of $A \Box\!\!\rightarrow C$, then CEM follows by elementary logic:

[1] The operators ϕ and ψ are said to be 'duals' when $\phi(\neg X)$ is equivalent to $\neg\psi(X)$. For example, $\exists$ and $\forall$ are duals: $\forall x \neg Fx$ is equivalent to $\neg\exists x Fx$.

$(A \mathrel{\Box\!\!\rightarrow} C) \equiv \neg\neg(A \mathrel{\Box\!\!\rightarrow} C)$ (Logical truth)
$(A \mathrel{\Box\!\!\rightarrow} C) \equiv \neg(A \mathrel{\Box\!\!\rightarrow} \neg C)$ (Since $A \mathrel{\Box\!\!\rightarrow} \neg C$ is equivalent to $\neg(A \mathrel{\Box\!\!\rightarrow} C)$
$\therefore (A \mathrel{\Box\!\!\rightarrow} C) \vee (A \mathrel{\Box\!\!\rightarrow} \neg C)$ (Logic)

See Williams, p. 691. But Lewis could respond by pointing out that while $A \mathrel{\Box\!\!\rightarrow} \neg C$ may not be the negation of $A \mathrel{\Box\!\!\rightarrow} C$, the two are indeed inconsistent. All this requires is the Law of Counterfactual Non-Contradiction:

$\neg((A \mathrel{\Box\!\!\rightarrow} C) \,\&\, (A \mathrel{\Box\!\!\rightarrow} \neg C))$

Given CNC, which Lewis's semantics imply (provided we ignore any case where A is impossible), $A \mathrel{\Box\!\!\rightarrow} \neg C$ is the contrary (though not the contradictory) of $A \mathrel{\Box\!\!\rightarrow} C$, and so asserting either is a way of denying the other.

(c) Williams (p. 692) discusses an argument from von Fintel and Iatridou that for all values of F and G:

$[\forall x : Fx]\neg(Fx \mathrel{\Box\!\!\rightarrow} Gx) \equiv [\forall x : Fx](Fx \mathrel{\Box\!\!\rightarrow} \neg Gx)$

Take a look at the argument and see what you think. If it is correct, then CEM holds for any counterfactual bound by a universal quantifier.

So while the case for CEM is not non-existent, in my opinion it's not strong enough to provide a independent reason to deny DT. But DT faces problems that have nothing directly to do with CEM.

2. Bennett, following Stalnaker, points out that DT treats 'might'-counterfactuals as idioms: their meaning doesn't come from the ordinary meanings of 'If…' and 'might' but is to be understood "as a single linguistic lump" (p. 191).

3. In general, statements of the form 'P, but I think that Q' are unassertable when P and Q are inconsistent. For example, 'It's raining, but I think that it's not raining' is something you just can't say. So the existence of acceptable assertions of the form 'P, but I think that Q' is good evidence that P is not inconsistent with Q.

   Now consider:

   > "If I had left the house with only five minutes to spare, I might have missed the train; still, I think that if I'd left the house then, I wouldn't have missed it."

   Though it's a bit clunky to include the antecedent again in the second conjunct, it's easy to imagine circumstances in which that statement would be a fine thing to assert. That's good evidence that P is not inconsistent with Q:

   P If I had left the house with only five minutes to spare, I might have missed the train.
   Q If I had left the house with only five minutes to spare, I would not have missed the train.

   But given DT, P *is* inconsistent with Q; indeed, P is simply equivalent to the negation of Q. More generally, DT predicts that assertions of '$A \mathrel{\Diamond\!\!\rightarrow} C$, but I think that $A \mathrel{\Box\!\!\rightarrow} \neg C$' are just instances of 'P, but I think that $\neg P$', and should sound just as bad as the rest. If you think (as I do) that there are many conceivable instances of the first conjunction that don't sound bad at all, you'll take this to be strong evidence against DT. (See Stalnaker (1978, p. 100) for his version of this point; he credits Richard Thomason with the suggestion.)

4. Williams (pp. 693-694) makes the following point against DT. On Tuesday, I consider the possibility that I will flip this fair coin tomorrow. "If I were to flip it tomorrow, it might land heads" and "If I were to flip it tomorrow, it might land tails" both seem true. But suppose that the world is deterministic, and it is now true that I will flip it tomorrow and it will land heads. Lewis and most others accept that $A \,\&\, C$ implies $A \mathrel{\Box\!\!\rightarrow} C$. (We will talk more about this next week, but for now you should see that it is plausible.) That means that "If I were to flip it tomorrow, it would land heads" is now true, and hence given DT, "If I were to flip it tomorrow, it might land tails" is now false.

   A DT-er could respond to this by denying that $A \,\&\, C$ implies $A \mathrel{\Box\!\!\rightarrow} C$; that's a high price. She could also bite the bullet: if the world is deterministic (she could say), then given our epistemic situation vis-a-vis the future, lots of 'might'-counterfactuals that seem true to us are, in fact, false, and we would regard them as such if we were fully informed of the relevant deterministic facts. A third way to respond is to deny that future-directed subjunctive conditionals express counterfactuals. It's weird to treat a statement beginning with "If I were to flip it…" as a counterfactual when no one is presupposing that the speaker won't flip the coin. Since we are considering DT as a truth about 'might' *counterfactuals*, this would make the example irrelevant.

5. DT underwrites an argument form that can be used to show that very many apparently true contingent counterfactuals are, in fact, false. We'll consider this closely in a bit.

ALTERNATIVES TO DT

1. Stalnaker defines $\Diamond\!\!\rightarrow$ as follows:

   ST $A \Diamond\!\!\rightarrow C =_{df.} \Diamond(A \Box\!\!\rightarrow C)$

   What kind of possibility is expressed by the $\Diamond$ on the right-hand side? Stalnaker's answer: whatever kind of possibility 'might' generally expresses. If 'might' can be used to express different kinds of possibility, so can $\Diamond\!\!\rightarrow$. Stalnaker thinks it can; he writes, "I think the most common kind of possibility which this word is used to express is epistemic possibility [...] But *might* sometimes expresses some kind of non-epistemic possibility" (1978, 99).

2. Lewis argues against ST with the following case. Suppose that I didn't look in my pocket and there is no penny in my pocket. Lewis's bold assertion: "Then 'If I had looked, I might have found a penny' is plainly false" (*Counterfactuals*, p. 80). If the $\Diamond$ in ST expresses logical or metaphysical possibility, then ST implies that the 'Looked $\Diamond\!\!\rightarrow$ Found' is true (since it's of course logically or metaphysically possible that 'Looked $\Box\!\!\rightarrow$ Found' should be true). Lewis's criticism was aimed at this version of ST (he was writing before Stalnaker had published a theory of $\Diamond\!\!\rightarrow$). But even if we read $\Diamond$ in ST as expressing epistemic possibility, ST could easily render a 'true' verdict for 'Looked $\Diamond\!\!\rightarrow$ Found': if it's epistemically possible that I have a penny in my pocket, then (ignoring weird counterexamples) it's epistemically possible that Looked $\Box\!\!\rightarrow$ Found.

   To me, this argument is evidence that Lewis, for all his brilliance, had a tin ear. I do not find 'Looked $\Diamond\!\!\rightarrow$ Found' to be anything like "plainly false". It seems to me that it could be just the thing to say. I don't know whether or not there are any pennies in my pocket. You instruct me: "Without checking your pockets, answer the following question: if you had looked in your pocket a moment ago, would you have found a penny there? Please answer using a conditional." The *obviously* correct answer: "Well, Looked $\Diamond\!\!\rightarrow$ Found, and Looked $\Diamond\!\!\rightarrow \neg$ Found."[2]

3. Bennett travels down a rabbit hole chasing after a bad suggestion from Stalnaker. The suggestion is that the $\Diamond$ in ST concerns "what would be compatible with [my knowledge] if I knew all the relevant facts". With this weirdly idealized notion of epistemic possibility (or "quasi-epistemic possibility") in hand, Bennett (pp. 191-192) argues:

   1. There are cases where ST misfires when we treat $\Diamond$ in this weird way (the coin-tossing case on p. 191).
   2. It's better to render $A \Diamond\!\!\rightarrow C$ as $A \Box\!\!\rightarrow \Diamond C$ (with $\Diamond$ being taken in the weird way).
   3. Finally, even though DT is not a definition of $\Diamond\!\!\rightarrow$, it is true, because given the weird understanding of $\Diamond$, $A \Box\!\!\rightarrow \Diamond C$ should be true just when $\neg(A \Box\!\!\rightarrow \neg C)$ is true.

   In §5 of his paper, Williams argues against number 3, which he calls "Bennett's Hypothesis". My qualms begin earlier, with the weird idea about $\Diamond$. I confess to not really understanding it; or, to the extent that I do, not seeing how there is any sense (even 'quasi') in which it qualifies as a kind of epistemic possibility. See the paragraph at the bottom of p. 190 if you are interested in trying to figure out what he's talking about.

   More broadly, I think Bennett's discussion suffers from an impoverished understanding of epistemic possibility. He treats it as mere compatibility with what is known. But this is too weak; e.g., no necessary falsehood is compatible with what I know, but many necessary falsehoods are epistemic possibilities for me. Or, $\neg P$ can be epistemically possible for me even when I know something that entails P, provided I haven't yet figured that out.

4. DeRose offers a restricted version of ST, suggesting that $A \Diamond\!\!\rightarrow C$ is equivalent to 'It is epistemically possible that $A \Box\!\!\rightarrow C$', where epistemic possibilities are identified as "possibilities of the kind that sentences of the form 'It's possible that P_{ind}' typically express" (DeRose 1999, 389). The '*ind*' subscript indicates that P is "in the indicative mood." The relevant contrast is best brought out with an example:

   ind It's possible that I do not exist.

   sub It's possible that I should not have existed.

   ind and *sub* seem clearly to express different kinds of possibilities. DeRose says the kind expressed by *ind* is epistemic possibility, and the kind expressed by *sub* is something else.

[2] Note that given DT, the second conjunct implies $\neg$(Looked $\Box\!\!\rightarrow$ Found). But certainly we wouldn't follow up by saying, "So the answer to your question is *no*." More generally: $A \Diamond\!\!\rightarrow C$ can be appropriate even when $\neg(A \Box\!\!\rightarrow \neg C)$ is not; given DT, this is surprising, since they are equivalent.

Given DeRose's suggestion (which he calls the 'Epistemic Thesis', or ET), we should expect just what the defender of DT says (see above): that instances of '$A \mathrel{\Box\!\!\rightarrow} C$ and $A \mathrel{\Diamond\!\!\rightarrow} \neg C$' will always sound horrible. They are equivalent to '$A \mathrel{\Box\!\!\rightarrow} C$ and it's epistemically possible that $A \mathrel{\Box\!\!\rightarrow} \neg C$' and so are instances of the Moore-paradoxical sentence form 'P and it is possible that $\neg P_{ind}$'; e.g., "I exist and it's possible that I don't exist." DeRose argues that since such conjunctions of would and might counterfactuals *always* sound horrible, we have reason to favor ET over ST, since on the latter view such conjunctions should sometimes sound fine (i.e., whenever the $\Diamond$ expresses non-epistemic possibility). That's because expressions of non-epistemic possibilities (e.g., sentences like *sub* above) don't produce Moore-paradoxical conjunctions: "I exist and it's possible that I should not have existed" is perfectly fine.

COUNTERFACTUAL SKEPTICISM

1. DeRose's version. Three conditionals:

   (A) If I had tagged up, I would have scored the winning run.
   (B) If I had tagged up, I might have tripped, fallen, and been thrown out.
   (C) If I had tagged up, I might not have scored a winning run.

   We start out certain that (A). But then we are asked to agree with (B), and can't resist. (B) leads to (C). And (C) seems inconsistent with (A). According to DT, it is. So how are we to respond?

   Deny (A). This is the "skeptical" response.
   Deny (B). But come on: really?
   Accept (B) and deny (C). But come on: ***really***?

   DeRose asks: "if (A) isn't true, what counterfactual conditional *is* contingently true?" Whenever we're certain that $A \mathrel{\Box\!\!\rightarrow} C$, there's *some* $\neg C$ possibility lurking that will lead us to think that $A \mathrel{\Diamond\!\!\rightarrow} \neg C$, and (by analogous reasoning) will require us to deny $A \mathrel{\Box\!\!\rightarrow} C$. So accepting the skeptical response here should lead us to think that most contingent counterfactuals are false. DeRose's response is to deny DT and endorse ET. This enables him to say that (A) and (C) are compatible, thus resisting the skeptical response. In addition, ET explains the clash between (A) and (C): any sentence of the form 'P and it might be that $\neg P_{ind}$' exhibits the same phenomenon.

2. The quantum plate version. First appears (to my knowledge) in Hawthorne (2005):

   > On those interpretations of quantum mechanics according to which the wave function for a system delivers probabilities of location, it seems that in any mundane situation, there is always a small chance of some extremely bizarre course of events unfolding. Suppose I drop a plate. The wave function that describes the plate will reckon there to be a tiny [GP: *tiny!*] chance of the particles comprising that plate flying off sideways (Hawthorne 2005, 396).

   Luckily, I didn't drop the plate. Now consider:

   (A) If I had dropped the plate, it would have fallen to the floor.
   (B) If I had dropped the plate, it might have flown off sideways.
   (C) If I had dropped the plate, it might not have fallen to the floor.

   Ordinarily I would be certain of (A). But in a quantum state of mind, I'll be inclined to accept (B), and hence (C). So must I reject (A)? If the answer is "yes" then similar reasoning will lead us to reject most contingent counterfactuals.

3. So DT seems to present us with a stark choice: either accept that most ordinary counterfactuals are false, or deny the relevant 'might' counterfactuals. Neither is appealing. Since DT leads to this bad choice, it should be rejected.

4. Here's a suggestion for an alternative solution for a DT-er. Assume Lewis's semantics, so (A) and (C) are indeed inconsistent as DT implies. Suppose that the similarity ordering relevant to the truth of $A \mathrel{\Box\!\!\rightarrow} C$ shifts with context. In an ordinary context where (A) seems true, the similarity ordering has C true at all the closest A-worlds, and so (A) is true. But when we accept (B), the similarity ordering shifts: now some of the closest A-worlds are $\neg C$-worlds. That leads us to accept (C), and with the new ordering in place, (A) is false. This wouldn't imply that (A) is *always* or even *ordinarily* false, only that it's false relative to contexts where (C) is true.