Handout 8: Bennett, Chapter 10

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

TERMINOLOGY

1. Chapters 10-18 concern subjunctive conditionals, which Bennett distinguishes from indicatives on the basis of a grammatical feature (see Bennett §4-5; my Handout 1). I will generally refer to the conditionals with which he is concerned as *counterfactuals*. A counterfactual says something about what would be the case if *A* were the case, presupposing that *A* is not actually the case. It is not obvious that all subjunctives express counterfactuals. (Does "If you were to take the train, you'd have a very relaxing trip" seem like a counterfactual to you?)

2. Bennett uses '>' for the subjunctive conditional. Lewis used '$\Box\!\!\rightarrow$' for the 'were'-'would' counterfactual. Both symbols are potentially misleading (Bennett's because it also means 'greater than' and Lewis's for a reason we'll see next week). I'll generally use $\Box\!\!\rightarrow$, except when it's important to distinguish subjunctives from counterfactuals.

§63-64. POSSIBLE WORLDS

1. The approach we'll be exploring makes central use of *possible worlds*. The skeleton of the theory is:

   $A \Box\!\!\rightarrow C$ is true at α (i.e., the actual world) iff at the possible world(s) closest to α where A is true, C is true.

   Most of the interesting questions about the theory concern the '(s)' after 'world', and what 'closest' means. Before turning to them, we need to say something about possible worlds.

2. 'Possible world' entered the contemporary philosophical lexicon by way of modal logic in the first half of the twentieth century. Modal logic is what results when you add the operators $\Box$ and $\Diamond$ to a logical language:

   $\Box P \approx$ It is necessary that P.

   $\Diamond P \approx$ It is possible that P.

   $\Box P$ and $\Diamond P$ are standardly treated as duals and hence each is definable from the other:

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

   In the formal context, a 'possible world' is nothing more than a particular element of a model for a formal language containing $\Box$ and $\Diamond$. We can illustrate its role by adding $\Box$ to a simple propositional language $\mathcal{L}$:

   $\mathcal{L}$'s vocabulary consists of a stock of propositional letters, $\neg$, $\supset$, and $\Box$. $\mathcal{L}$'s formation rules are:

   If P is a propositional letter, then P is a formula.
   If P is a formula, then $\neg P$ is a formula.
   If P and Q are formulas, then $P \supset Q$ is a formula.
   If P is a formula, then $\Box P$ is a formula.

   A model $\mathcal{M}$ for $\mathcal{L}$ is a pair $\langle \mathbf{W}, \mathbf{V} \rangle$ such that: $\mathbf{W}$ is a set of objects, and $\mathbf{V}$ is an assignment of T or F to each pair $\langle \mathcal{A}, w \rangle$, where $\mathcal{A}$ is a propositional letter in $\mathcal{L}$ and $w \in \mathbf{W}$. $\mathbf{W}$ is the model's set of 'worlds', and $\mathbf{V}$ tells us which atomic propositions are T and which are F at each 'world'. Truth in $\mathcal{M}$ is characterized as follows:

   An atomic formula $\mathcal{A}$ is true at w in $\mathcal{M}$ iff $\mathbf{V}\langle \mathcal{A}, w \rangle$ = T.

$\neg P$ is true at w in $\mathcal{M}$ iff P is not true at w in $\mathcal{M}$.

$P \supset Q$ is true at w in $\mathcal{M}$ iff $\neg P$ is true at w in $\mathcal{M}$ or Q is true at w in $\mathcal{M}$.

$\Box P$ is true at w in $\mathcal{M}$ iff for all $w' \in \mathbf{W}$, P is true at w' in $\mathcal{M}$.

There are two ways to make the semantics for $\Box$ more complicated:

1. The clause for $\Box P$ above implies that if $\Box P$ is true at any world in $\mathbf{W}$, it is true at every world in $\mathbf{W}$. Given the definition of $\Diamond$ as the dual of $\Box$, the same holds for $\Diamond P$. For some kinds of necessity and possibility, these features are undesirable. Consider, for example, causal necessity: intuitively, P is causally necessary at w iff P is true at all worlds with w's causal laws; possible worlds where w's causal laws do not hold are irrelevant to causal necessity at w. To represent "weaker" kinds of necessity we add an *accessibility relation* to $\mathcal{M}$, and modify the clause for $\Box P$ to say that $\Box P$ is true at w in $\mathcal{M}$ iff for all $w' \in \mathbf{W}$ *accessible from* w, P is true at w' in $\mathcal{M}$. This allows us to have $\Box P$ and $\Diamond P$ true at some worlds without being true at all worlds.
2. Many interesting and difficult technical and philosophical questions emerge when we add $\Box$ to a language containing quantifiers, variables, constants, and identity.

But neither complication alters the basic formal role of a possible world: it is simply an index for one complete assignment of the values T or F to all of the atomic formulas of the language. It does not matter what the objects in $\mathbf{W}$ are. They can be the real numbers, regions of space, whatever—just so long as we have enough of them to index all of the assignments we are interested in.

3. The informal use of 'possible world' is directly analogous: it is a shorthand for talking about some way things are or could have been. My shoes are brown, but they could have been black instead. Translated into possible worlds talk, this banal observation becomes: at the actual world, my shoes are brown, but there is a different possible world at which they are black. Another truth, somewhat less banal and perhaps controversial: my shoes couldn't have been made out of argon. Translated: there's no possible world at which my shoes are made of argon. Equivalently: at all possible worlds, if my shoes exist at all they are made of something other than argon.

4. Some philosophers have taken this shorthand literally, or as an informative analysis of what we mean when we talk about what could or couldn't be. For them the metaphysical question, "What is a possible world?" is important. David Lewis infamously defended the bizarre (but strangely beautiful) answer: all possible worlds are, as Bennett puts it, "metaphysically on a par with" the actual world. For Lewis, the uncountably many possible worlds are all existing concrete realities of which ours is but one. We use the term "actual" to denote our world, and denizens of the other worlds use "actual" to denote theirs, but none is any more real than the others. In §63 Bennett calls this idea "extreme realism" and rehearses some standard arguments against it.

5. In §64 Bennett gives his own answer: possible worlds are "maximal consistent propositions." He treats "proposition" as interchangeable with "state of affairs" or "way things could be". I think "way things could be" is the intuitively basic concept in this neighborhood. It is somewhat less awkward to use "proposition," though I think that term has problematic connotations that "way things could be" does not. I do not think that "state of affairs" is apt, since I regard states of affairs as facts, and a fact is a way things are, not (merely) a way things could be. "Maximal" means that for any proposition P, a possible world either entails P or entails $\neg P$. "Consistent" means what it sounds like: no possible world entails a contradiction. Be careful: while a proposition's being maximal and consistent is necessary for it to be a possible world, it may not be sufficient. It is easy to suppose that there is a maximal consistent proposition that entails that there is a pair of argon shoes. Still, we may wish to deny that there could be argon shoes. What could and could not be seems not to be merely a matter of logical consistency.

6. Bennett calls his view "abstract realism": possible worlds are *real*, and they are *abstract*. My own understanding of possible worlds talk makes me an abstract realist in Bennett's sense, though both aspects of the label are unnecessarily provocative. I deny that possible worlds are concrete objects, so if 'abstract' and 'concrete' are jointly exhaustive, then possible worlds are abstract. And I think that what could and couldn't be is independent from my (or anyone's) conception of the world, and so I am a "realist" about possible worlds. Those in the grip of an all-too-common mindset will be eager to draw facile inferences from this: "Ah, so you admit abstract, mind-independent 'ways' into your ontology, do you? Well, what are you going to say about the nature and existence of these mysterious objects?" Do not be tricked into thinking that responsibly using possible world talk requires you first to

answer such questions. That things are a certain way but could have been other ways instead is a virtually indubitable platitude, and is made no less plausible by being restated with quantifiers ranging over ways or in terms of possible worlds. Metaphysical questions about the nature, existence, and variety of possible worlds or ways things could be will not concern us here, and we will have no qualms about our inability to give deep answers to them.

7. Dangerous confusion warning: Bennett points out, rightly, that for the abstract realist 'the actual world', when used in the relevant sense, refers not to the concrete reality around us, but to the way everything actually is (or, in Bennett's preferred terminology, the maximal proposition representing actuality). The actual world is a possible world, so if possible worlds are ways things could be (or maximal consistent propositions), the actual world is a way things could be (or a maximal consistent proposition), and hence isn't composed of rocks, trees, stars, etc. But what Bennett does not make clear is that this is a ***stipulative use*** of the term 'the actual world,' and not a proposal about what 'the actual world' normally means! Accepting abstract realism doesn't require accepting that rocks and stars are not part of the actual world (now employing that term in its ordinary sense).

8. Dangerous falsehood warning: on p. 158 Bennett writes: "with 'world' thus understood [i.e., as a maximal consistent proposition], we do not know which world we exist at [...] The whole reality of an abstract world is that of a certain *maximal* proposition. To know which such world is actual, then, would involve knowing which maximal proposition is actually true, which would be knowing the whole contingent truth." But knowing which maximal proposition is actually true is ***not*** the same as knowing the whole contingent truth. I can know which car is mine even though there are many things I do not know about my car; similarly, I can know which maximal proposition is actual even though there are many things I do not know about it. How? Because I know that the actual world is the one that matches reality precisely. Do not think this is trivial; it suffices to distinguish the actual world from all other possible worlds in any conceivable circumstance. I see no reason to accept Bennett's bewildering skeptical claim.

§65-66. Two false logical principles

1. A ***strict conditional*** says that $A \supset C$ holds throughout some set of possible worlds; equivalently, that relative to some domain of worlds, $\Box(A \supset C)$. The following conditional relations are strict:

   Entailment. A entails C iff at all possible worlds, $A \supset C$.[1]

   Causal implication. A causally implies C at w iff at all possible worlds that conform to w's causal laws, $A \supset C$.

   Material implication. A materially implies C at w iff at w, $A \supset C$.

   In the case of entailment, the domain is ***all possible worlds***; for causal implication, it is ***all possible worlds governed by w's causal laws***; for material implication it is ***all possible worlds identical with w***. We use $\strictif$ to denote strict implication.

2. Antecedent Strengthening and Transitivity are valid for any variety of $\strictif$:

   Antecedent strengthening. If $A \strictif C$, then $(A \,\&\, B) \strictif C$.

   Transitivity. If $A \strictif B$ and $B \strictif C$, then $A \strictif C$.

   As Bennett shows, Transitivity implies Antecedent Strengthening:

   If $(A \,\&\, B) \strictif A$ and $A \strictif C$, then $(A \,\&\, B) \strictif C$. (Transitivity)
   $(A \,\&\, B) \strictif A$ is a logical truth.
   So, if $A \strictif C$, then $(A \,\&\, B) \strictif C$.

   So if Antecedent Strengthening is invalid for a certain conditional relation, so is Transitivity.

3. Antecedent Strengthening is clearly invalid for counterfactual conditionals:

   ✔ If you had walked on the ice, the ice would have broken.

   × If you had walked on the ice while leaning heavily on the extended arm of someone standing on the shore, the ice would have broken.[2]

[1]So says Bennett on p. 159; I do not think this is an accurate characterization of entailment, but we'll let it pass.
[2]Lewis (*Counterfactuals*, p. 10) credits J. Howard Sobel with the discovery of such pairs, and so they are often called "Sobel sequences".

Given that Transitivity implies Antecedent Strengthening, the former is invalid for counterfactual conditionals, too. However, it is harder to confirm this with intuitive examples. Have a look at Lewis's counterexamples on pp. 32-33, then try and see if you can come up with any of your own. Bennett discusses two attempts to rescue Antecedent Strengthening for counterfactuals in §66. One is worth discussing, to see that it is a no-hoper:

4. Some may be tempted to treat ✔ above as elliptical for:

   ✔* If you had walked on the ice *and weren't holding on to anything*, the ice would have broken.

   Since × can't be reached from ✔* by Antecedent Strengthening, the pair is not a counterexample to either logical principle. But now consider ×*:

   ×* If you had walked on the ice and weren't holding on to anything and there was a steel walkway just below the ice's surface, the ice would have broken.

   Let your imagination run free, and you'll see how easy it is to continue this game. The only way to bring it to an end is to say that ✔ is elliptical for:

   ✔^{T} If you had walked on the ice *and nothing was the case that would prevent your doing so from breaking the ice*, the ice would have broken.

   But unlike ✔, which is a very useful thing to know and could very well be false, ✔^{T} is a boring triviality that no one would deny. Antecedent Strengthening can be rescued, but only by sapping the life from counterfactuals.

§67. VARIABLY STRICT CONDITIONALS

1. Lewis's idea is that counterfactuals are "variably strict". Bennett's description of this elegant idea is needlessly confusing, so I will give my own. See Lewis's *Counterfactuals* §1.3 for a formally rigorous development.

2. The basic idea is that $A \Box\!\!\rightarrow C$ determines a set of possible worlds, and then states that $A \supset C$ is true throughout that set; equivalently, that $A \strictif C$ is true with respect to the determined set of worlds. This conditionals are *variably* strict because different counterfactuals determine different sets of worlds. If $A \Box\!\!\rightarrow C$ and $(A \,\&\, B) \Box\!\!\rightarrow C$ determine different sets of worlds, we should not be surprised that Antecedent Strengthening fails for $\Box\!\!\rightarrow$. So what set of worlds is determined by $A \Box\!\!\rightarrow C$?

3. Picture a set of nested spheres. At the center lies the actual world (α). On each sphere around α lie other possible worlds. The spheres are ordered by the "comparative overall similarity" of the worlds on that sphere to α. (This is an initial stipulation to get the theory off the ground: as we'll see, there is quite a bit more to say about the ordering.) If w is on sphere S, then all worlds on spheres inside S are more similar to α than w, all worlds on spheres outside S are less similar to α than w, and any other worlds on S are just as similar to α as w. Lewis calls this picture a "Ptolemaic astronomy" of worlds (Bennett reports that the label was first used by Mackie as a term of derision).

4. Now consider the smallest sphere around α containing worlds where A is true ('A worlds'). $A \Box\!\!\rightarrow C$ says that all worlds on that sphere are C worlds; equivalently, that $A \supset C$ is true at all worlds on or within that sphere; equivalently again, that $A \strictif C$ is true with respect to the set of all worlds on or within that sphere; equivalently again, that no $A \,\&\, \neg C$ world is closer to α than any $A \,\&\, C$ world.

5. Applied to the conditionals above:

   ✔ "If you had walked on the ice, the ice would have broken" is true at α iff at all the possible worlds where you walked on the ice closest to α, the ice broke.

   × "If you had walked on the ice while leaning heavily on the extended arm of someone standing on the shore, the ice would have broken" is true at α iff at all the possible worlds where you walked on the ice leaning heavily on the extended arm of someone standing on the shore, the ice broke.

   It is easy to see that the condition for ” can obtain without the condition for × obtaining.

6. Two consequences of the account:

   (a) If A is true at α, then $A \mathrel{\Box\!\!\rightarrow} C$ is true at α iff C is true at α; i.e., $A \mathrel{\Box\!\!\rightarrow} C$ collapses into $A \supset C$.

   (b) If A is not true at any possible world, then $A \mathrel{\Box\!\!\rightarrow} C$ is trivially true: since there are no A worlds, no $A \,\&\, \neg C$ world is closer to α than any $A \,\&\, C$ world. (One could resist this consequence by allowing spheres of impossible worlds, but it would not be easy to order them by their overall similarity to α!)

7. Note that since possible worlds are simply ways things could have been, we can restate the theory as: $A \mathrel{\Box\!\!\rightarrow} C$ is true iff all of the ways things could have been in which A would have been true most similar to the way things actually are are also ways things could have been in which C would have been true. Though cumbersome, that makes perfectly good sense.

§68. CONDITIONALS WITH DISJUNCTIVE ANTECEDENTS

1. Simplification of Disjunctive Antecedents is invalid if the worlds theory is correct:

   SDA If $(A \vee B) \mathrel{\Box\!\!\rightarrow} C$, then $A \mathrel{\Box\!\!\rightarrow} C$ and $B \mathrel{\Box\!\!\rightarrow} C$.

   Suppose there are some A-worlds closer to α than any B worlds, and that those closest A-worlds are all C worlds. Then the closest $A \vee B$ worlds are C worlds. So $(A \vee B) \mathrel{\Box\!\!\rightarrow} C$ is true. But all of this is compatible with the closest B-worlds not being C-worlds, and hence $B \mathrel{\Box\!\!\rightarrow} C$'s being false. Thus SDA is invalid.

2. There is intuitive evidence for SDA. Bennett's example:

   D If there had been rain or frost, the game would have been called off.

   D_r If there had been rain, the game would have been called off.

   D_f If there had been frost, the game would have been called off.

   D certainly seems to imply both D_r and D_f. How can we square this with SDA's invalidity?

3. Bennett considers two answers:

   1. Someone who asserts D is actually asserting $D_r \,\&\, D_f$.

      Bennett says (and I agree) that this is is ad hoc.

   2. For Gricean reasons, it's appropriate to assert D only if you are also willing to assert D_r and D_f. If you don't think D_f is true, then it would be misleading but might not be false for you to assert D. Thus we mistake an implicature of an assertion of D (to wit, $D_r \,\&\, D_f$) for an entailment. I would like for this to be right, but I haven't quite been able to see how the Gricean explanation is supposed to go.