# Conditionals

Phil 296 - November 16

Start with three questions:

1. What do ordinary utterances of the form "If this had happened, that would have happened" mean? When are they true?
2. Given the role that sentences like those play in a lot of important tasks, e.g, assigning legal responsibility, planning complex actions, evaluating past actions, explaining and understand events, what do those sentences mean in those particular legal/political/scientific contexts?
3. What are some simple mathematical structures that yield accounts of what these sentences mean that are, at least roughly, like the answers to 1 and 2?

And once we have an answer to 3, we can ask two more questions.

4. How can we tinker with those models to make them even better answers to 1 and 2.
5. Can we use the models to resolve puzzle cases for 1 and 2 that we couldn't previously answer?

This is a logic class, so we're going to mostly care about whether the structures are 'at least roughly' right in terms of which arguments should or should not be valid.

## Model 1 - Material Implication

If $A$ were true, $B$ would be true just means $A \supset B$. This leads to absurd results. Think about the meme that is "This is the world we would have had if …".

On the material implication account, the implicit conditional there is true every single time. After all, in the meme versions, the $A$ is always false, and $A \supset B$ is true whenever $A$ is false. Since that is not the world we would have had if, for example, we'd encouraged everyone to get COVID as quickly as possible, this theory is false.

## Model 2 - Strict Implication

This is more plausible. It says that If $A$ were true, $B$ would be true just means $\Box(A \supset B)$. This is sometimes written with a fishhook between $A$ and $B$, but I don't have that font on my computer, so I won't write it here!

First problem, this seems to make basically all of these conditionals come out false. It seems true that if I'd saved more money, I'd have more money now. But there's a possible world where there is a malevolent demon who would have incinerated my money if I'd saved more.

Well maybe we can solve that by saying one of the two following things.

- The $\Box$ here is connected to a modal logic where the accessible worlds are the 'normal' worlds, worlds that don't have demons, etc.
- What the conditional really says is $\Box((A \wedge N) \supset B)$, where N means that things are normal.

I won't repeat the proof here, but these approaches turn out to be equivalent.

This view implies that the following three implications are always valid. (Note I'll use $\rightarrow$ here for the English 'if … then', and all other symbols will be bits of mathematical notation that should get

**Antecedent Strengthening** $A \rightarrow B \vdash (A \wedge C) \rightarrow B$
**Transitivity** $A \rightarrow B, B \rightarrow C \vdash A \rightarrow C$
**Contraposition** $A \rightarrow B \vdash \neg B \rightarrow \neg A$

And there are a lot of cases in English where these all seem to fail. Note that the first two are connected. Assume that all instances of transitivity are correct. Then the following argument is going to be good.

- $(A \wedge C) \rightarrow A, A \rightarrow B \vdash (A \wedge C) \rightarrow B$

But since the first premise is (surely!) a logical truth, we can drop it, so this argument is valid.

- $A \rightarrow B \vdash (A \wedge C) \rightarrow B$

And that's just antecedent strengthening.

## Model 3 - Normal A-worlds

Stipulate that we have a new function $N$ from propositions to propositions. So if $A$ is a proposition, so is $N(A)$. Intuitively, $N(A)$ means that $A$ is true in a 'normal' way, but we'll have to build that intuition into the definition of $N$ somehow.

Hypothesis: If $A$ were true, $B$ would be true just means $\Box(N(A) \supset B)$. So it means that B is true in all normal A-worlds.

Good news: This means that the three problem inferences: Antecedent Strengthening, Transitivity, and Contraposition all fail.

Bad news: Literally nothing can be shown to hold so far, since we've said literally nothing about $N$. (Well, that's a bit strong, since $\Box B \vdash A \rightarrow B$ is valid, but that's about it.)

Move: Tinker with the definition of $N$ so that it does the job we want. Since $N$ lets in too much right now, it lets for example the normal $A$-worlds be the worlds where Barack Obama teaches this class, the obvious move is to add **constraints** on $N$, restricting what $N$ could be mathematically.

To express these constraints, we'll do explicitly something I've been doing implicitly a lot so far, identifying propositions with the sets of worlds where they are true.

First constraint: $N(A) \subseteq A$. That is, all worlds where $N(A)$ holds are worlds where $A$ holds. This was originally part of the intuitive conception, but we hadn't fixed it. This is sufficient (and kind of necessary) to guarantee that $A \rightarrow A$ is a logical truth.

Second constraint: Let @ be a special name for the actual world. The constraint is if $@ \in A$, then $@ \in N(A)$. If the actual world is an $A$-world, it is a normal $A$-world.

I mentioned in class a strengthening of that constraint: if $@ \in A$, then $\{@\} = N(A)$. That is, if the actual world is an $A$-world, then it is the **only** normal $A$-world. How would things be if Michigan had won last Saturday? They would be exactly like they are, in **every** respect. We'll come back to this constraint later (probably next week).

This gives us a logic, but it's still a very weak one. Note that it puts no constraints on how normality relates different propositions. Imagine that all the following are true.

- $A \to B$, i.e., the normal $A$-worlds are $B$-worlds.
- $B \to A$, i.e., the normal $B$-worlds are $A$-worlds.
- $A \to \neg C$, i.e., no normal $A$-worlds are $C$-worlds.
- $B \to \neg C$, i.e., no normal $B$-worlds are $C$-worlds.

It feels like that should be enough to say that $(A \wedge B) \to \neg C$. The normal $A \wedge B$ worlds just are some of those normal $A$-worlds and $B$-worlds. But nothing we've said so far puts any constraints whatsoever on how $N$ behaves with respect to different inputs. So actually these are all consistent with $(A \wedge B) \to \neg C$ being false, and even with $(A \wedge B) \to C$ being true.

The constraints that Priest numbers 3 through 5 are designed to solve problems like this one, but by this stage I think the idea that we have any kind of simple, elegant, model has gone out the window. So let's try a different approach.

## Model 4 - Distances

Assume that there is a function $d$ from pairs of worlds to distances. For now assume distances are non-negative integers. (We'll come back to this assumption maybe.) So read $d(w_1, w_2) = n$ as the distance between $w_1$ and $w_2$ is $n$.

Say that a **sphere** of worlds around $w$ is a set consisting of all and only the worlds that are a particular distance from $w$. So sphere $S_1$ is the set of all worlds that are distance 1 from $w$. For any world there will be a different set of spheres centered on it.

Here's the new proposed truth conditions for $A \to B$. They are disjunctive. The sentence is true, at $w$, if either of the following two conditions are met:

1. There are no worlds where $A$ is true; or
2. At the nearest sphere that contains any $A$-worlds, all $A$-worlds are $B$-worlds.

As the other handout shows, this approach yields counterexamples to Antecedent Strengthening, Transitivity, and Contraposition. But it makes some related things come out true.

**Modified Transitivity** A →B, (A  B) →C  A →C

This is important because, coming back to question 2, we really do seem to want something like transitivity when evaluating how policies worked. Imagine someone who argues that we should have had stricter mask mandates in 2020 arguing as follows:

1. If we'd had stricter mandates, we'd have had fewer cases.
2. If we'd had fewer cases, fewer people would have died.
3. So if we'd had stricter mandates, fewer people would have died.

You can quibble on the premises - there is still a big debate about whether mandates work and so whether premise 1 is true - but the reasoning seems kind of ok. It would be weird for a mandate opponent to say, yes those premises are fine, but they don't support your conclusion. So what's gone on here? One hypothesis is that both supporters and opponents of the argument are really reaading the second premise as this:

- If we'd had **mask mandates and** fewer cases, fewer people would have died.

And now the argument is valid, and we can go back to what seems like the interesting policy question: whether premise 1 is actually true.

## Modifying Model 4

If you like this model so far, then you might want to tinker with it. The natural way to 'tinker' is to put constraints on $d$. Like with $N$, we introduced this as a function, and with an intended interpretation, but we didn't do the hard work of showing that the function really satisfied the intended interpretation. Just like with modal logic, and with $N$, the more constraints you put on $d$, in general the more things you can prove. Here are some constraints that seem natural.

- $d(w, w) = 0$

That is, every world is distance 0 from itself. Without that, we actually don't have $A \to B, A \vdash B$. It could be that the 'nearest' $A$-worlds do not include the actual world. This seems like a very natural constraint on $d$.

Here are two more constraints that make sense given the 'distance' metaphor, but as far as I know make no difference to the logic.

- $d(w_1, w_2) = d(w_2, w_1)$
- $d(w_1, w_2) + d(w_2, w_3) \leq d(w_1, w_3)$

A more interesting constraint, also suggested by the distance metaphor, is this one:

- If $w_1 \neq w2$, then $d(w_1, w_2) > 0$

That is, all distinct worlds have positive distance between them. If that's right, we get the following as a valid argument schema.

- $A, B \vdash A \to B$

Why? Assume $A, B$ are true at a particular world $w$. The sphere of nearest $A$-worlds will just be $\{w\}$, since it is distance 0 away, and all other worlds are positive distance away. And $B$ is true at all of the worlds in $S_0$.

## Conditional Excluded Middle

I don't think we'll get to this today, but I wanted to note that one of the big debates that comes next concerns the following two (equivalent) principles:

- $\vdash (A \to B) \vee (A \to \neg B)$
- $\vdash (A \to \neg B) \equiv \neg(A \to B)$

Should those be true? Actually, going back to our original question, we can separate those out - should those be true in ordinary language, and should they be true in the artificial language we use for formally modeling decisions like mask mandates? Maybe those get different answers.

Anyway, they don't hold in any of the last three models we looked at, so if we want to make them hold, we need to tinker some more. If we have time, we'll look at the tinkering that is needed.