

# 8. Conditionals

## 8.1. Material conditionals

A good thing about sentences of the form  $A \rightarrow B$  (called **material conditionals**) is that it is complete clear what they mean.  $A \rightarrow B$  is true iff  $A$  is false or  $B$  is true. It is not so clear whether ‘if  $A$  then  $B$ ’ sentences in natural language can be adequately translated as  $A \rightarrow B$ .

The following facts about logical consequence (in classical propositional logic) are widely thought to show that English conditionals are not material conditionals.

- (P1)  $B \models A \rightarrow B$
- (P2)  $\neg A \models A \rightarrow B$
- (P3)  $\neg(A \rightarrow B) \models A$
- (P4)  $A \rightarrow B \models \neg B \rightarrow \neg A$
- (P5)  $A \rightarrow B \models (A \wedge C) \rightarrow B$

(P1)–(P5) are sometimes called “paradoxes of material implication”, although they aren’t really paradoxical. But they do sound wrong if we assume that  $A \rightarrow B$  translates ‘if  $A$  then  $B$ ’ or (worse) ‘ $A$  implies  $B$ ’. Here are apparent counterexamples for each schema.

1. The lecture ends at 3pm. Therefore: If the building collapses at 2.45 then the lecture ends at 3pm.
2. The President won’t be impeached. Therefore: If the President will be impeached then nobody will care.
3. It is not the case that if it will rain tomorrow then the Moon will fall onto the Earth. Therefore: It will rain tomorrow.

4. If our opponents are cheating, we will never find out. Therefore: If we will find out that our opponents are cheating, then they aren't cheating.
5. If you add sugar to your coffee, it will taste good. Therefore: If you add sugar and vinegar to your coffee, it will taste good.

Almost everyone agrees that these inferences “sound wrong”. But not everyone agrees that they refute the hypothesis that English conditionals are (equivalent to) material conditionals. Some hold that what's wrong with the inferences is merely pragmatic: in a context in which it would be sensible to utter the premise, it would not be sensible to utter the conclusion. On that account, the inferences are actually valid, in the sense that there is no conceivable scenario in which the premises are true and the conclusion false.

There are also direct arguments in favour of the interpretation of English conditionals as material conditionals. For example, suppose I make the following promise.

- (1) If I don't have to work tomorrow, then I will help you move.

Under what conditions will I have broken my promise? I have made a false promise if the next day I don't have to work and yet I don't help you move. Under all other conditions, however, you couldn't fault me for having broken my promise. So it seems that (1) is false iff I don't have to work and I don't help you move. Generalizing, this suggests that ‘if  $A$  then  $B$ ’ is false iff  $A$  is true and  $B$  is false. So ‘if  $A$  then  $B$ ’ is true iff either  $A$  is false or  $B$  is true.

Another argument starts with the intuitively plausible assumption that ‘ $A$  or  $B$ ’ entails the corresponding conditional ‘if not- $A$  then  $B$ ’. (This is sometimes called the *or-to-if principle*.) For instance, if I tell you that Nadia is either in Rome or in Paris, you can infer that if she's not in Rome then she's in Paris. Now we can reason as follows.

1. Suppose ‘not- $A$  or  $B$ ’ is false.
2. Then  $A$  is true and  $B$  is false (by the meaning of ‘not’ and ‘or’).
3. Then the conditional ‘if  $A$  then  $B$ ’ is clearly false.
4. So the falsehood of ‘not- $A$  or  $B$ ’ entails the falsehood of ‘if  $A$  then  $B$ ’.
5. By the or-to-if principle, the truth of ‘not- $A$  or  $B$ ’ entails the truth of ‘if  $A$  then  $B$ ’.

## 8. Conditionals

---

6. So ‘if  $A$  then  $B$ ’ is logically equivalent to ‘not- $A$  or  $B$ ’.

Many more arguments have been given for and against the hypothesis that natural-language conditionals are material conditionals. We won’t look further into this debate. In any case, even those who defend the reading of English conditionals as material conditionals admit that it does not work for all if-then sentences in natural language.

One kind of counterexample are generic conditionals like (2).

(2) If water is heated to  $100^{\circ}$  C, it evaporates.

This shouldn’t be translated as  $p \rightarrow q$ . Intuitively, (2) states that *in all (normal) cases* where water is heated to  $100^{\circ}$  C, it evaporates. It is a quantified, or modal claim.

Another kind of counterexample are so-called subjunctive conditionals. Compare the following two statements.

(3) If Shakespeare didn’t write *Hamlet*, then someone else did.

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

(3) seems true. Someone has written *Hamlet*; if it wasn’t Shakespeare then it must have been someone else. But (4) is almost certainly false. After all, it is very likely that Shakespeare did write *Hamlet*. And it is highly unlikely that if he hadn’t written *Hamlet* – if he got distracted by other projects, say – then someone else would have stepped in to write the exact same piece.

Sentences like (3) are called **indicative conditionals**; (4) is a **subjunctive conditional**. (Subjunctive conditionals are also called ‘counterfactual conditionals’ or ‘counterfactuals’.)

Whatever we say about indicative conditionals like (3), subjunctive conditionals clearly can’t be translated as material conditionals. As I just said, on the assumption that Shakespeare wrote *Hamlet*, (4) is almost certainly false, even though it has a false antecedent, and so the corresponding material conditional is true.

To sum up, there seem to be different kinds of conditionals – different kinds of if-then sentences – in natural language. At least some of them can’t be translated as material conditionals. If we want to formalize reasoning with these conditionals, we need a better translation.

## 8.2. Strict conditionals

One apparent difference between material conditionals  $A \rightarrow B$  and conditionals in natural language is that  $A \rightarrow B$  requires no connection between the antecedent  $A$  and the consequent  $B$ . Consider (1).

- (1) If we leave after 5, we will miss the train.

Intuitively, if someone utters (1), they want to convey that missing the train is a *necessary consequence* of leaving after 5 – that it is impossible to leave after 5 and still make it to the train (given certain facts about the distance to the station, the time it takes to get there, etc.). This suggests that (1) should be formalized not as  $p \rightarrow q$  but as  $\Box(p \rightarrow q)$  or, equivalently,  $\neg\Diamond(p \wedge \neg q)$ .

Sentences of the form  $\Box(A \rightarrow B)$  or  $\neg\Diamond(A \wedge \neg B)$  are called **strict conditionals**. The label goes back to C.I. Lewis (1918), who also introduced the abbreviation  $A \dashv\vdash B$  for strict conditionals.

Lewis was not interested in the interpretation of ordinary-language conditionals. He wanted  $A \dashv\vdash B$  to formalize ‘ $A$  implies  $B$ ’ or ‘ $A$  entails  $B$ ’. His intended use of  $\dashv\vdash$  roughly matches our use of the double-barred turnstile ‘ $\models$ ’. But there are important differences. The turnstile is an operator in our *meta-language*; Lewis’s  $\dashv\vdash$  is an *object-language* operator like  $\wedge$  or  $\rightarrow$  that can be placed between any two sentences in a formal language to generate another sentence in the language. For example,  $p \dashv\vdash (q \dashv\vdash p)$  is well-formed, whereas  $p \models (q \models p)$  is gibberish. Moreover, while  $p \models q$  is simply false – because there are models in which  $p$  is true and  $q$  false – Lewis’s  $p \dashv\vdash q$  is true on some interpretation of the sentence letters and false on others. For instance, if  $p$  means that it is snowing and  $q$  that precipitation occurs, then  $p \dashv\vdash q$  is plausibly true, because snowfall is a form of precipitation, so the hypothesis that it is snowing implies that precipitation occurs.

Let’s set aside Lewis’s project of formalizing the intuitive concept of implication. Our goal is to define an object-language operator that functions like ‘if . . . then . . .’ in English. To see whether Lewis’s  $\dashv\vdash$  can do the job, we need to have a closer look at what it means.

Since  $A \dashv\vdash B$  is equivalent to  $\Box(A \rightarrow B)$ , standard Kripke semantics for the box also provides a semantics for strict conditionals. In Kripke semantics,  $\Box(A \rightarrow B)$  is true at a world  $w$  iff  $A \rightarrow B$  is true at all worlds  $v$  accessible from  $w$ . And  $A \rightarrow B$  is true at

## 8. Conditionals

---

$v$  iff either  $A$  is false at  $v$  or  $B$  is true at  $v$ . So we get the following truth-conditions for strict conditionals.

### Definition 8.1: Kripke semantics for $\rightarrow$

If  $M = \langle W, R, V \rangle$  is a Kripke model, then

$M, w \models A \rightarrow B$  iff for all  $v$  such that  $wRv$ , either  $M, v \not\models A$  or  $M, v \models B$ .

### Exercise 8.1

We can define  $\rightarrow$  in terms of  $\Box$  and  $\rightarrow$ . Can you define  $\Box$  in terms of  $\rightarrow$  and truth-functional operators? That is, can you find a sentence schema with  $\rightarrow$  as the only non-truth-functional operator that is equivalent (in Kripke semantics) to  $\Box A$ ?

As always, the logic of strict conditionals depends on what constraints we impose on the accessibility relation. For example, without any constraints,  $\rightarrow$  does not validate *modus ponens*, in the sense that  $A \rightarrow B$  and  $A$  together do not entail  $B$ . We can easily see this by translating  $A \rightarrow B$  back into  $\Box(A \rightarrow B)$  and setting up a tree. Recall that to test whether some premises entail a conclusion, we start the tree with the premises and the negated conclusion.

1.  $\Box(A \rightarrow B)$  (w) (Ass.)
2.  $A$  (w) (Ass.)
3.  $\neg B$  (w) (Ass.)

With the K-rules, where we don't make any assumptions about the accessibility relation, node 1 can't be expanded, so there is nothing we can do.

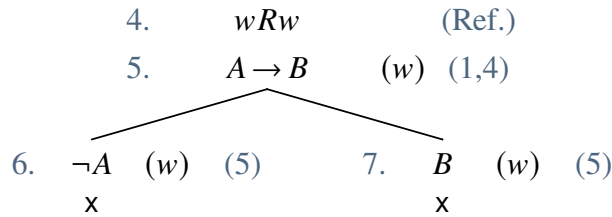
### Exercise 8.2

Give a countermodel in which  $p \rightarrow q$  and  $p$  are true at some world while  $q$  is false.

On the other hand, if we assume that the accessibility relation is reflexive, the tree closes:

## 8. Conditionals

---



It is not hard to show that *modus ponens* for  $\rightarrow$  is valid on all and only the reflexive frames. So reflexivity is precisely what we need to render *modus ponens* valid.

### Exercise 8.3

Confirm the following claims, by translating  $A \rightarrow B$  into  $\Box(A \rightarrow B)$ .

- (a)  $\models_K A \rightarrow A$
- (b)  $\neg B \rightarrow \neg A \models_K A \rightarrow B$
- (c)  $A \rightarrow B, B \rightarrow C \models_K A \rightarrow C$
- (d)  $(A \vee B) \rightarrow C \models_K (A \rightarrow C) \wedge (B \rightarrow C)$
- (e)  $A \rightarrow (B \rightarrow C) \models_T (A \wedge B) \rightarrow C$
- (f)  $A \rightarrow B \models_{S4} C \rightarrow (A \rightarrow B)$
- (g)  $((A \rightarrow B) \rightarrow C) \rightarrow (A \rightarrow B) \models_{S5} A \rightarrow B$

If we want  $A \rightarrow B$  to translate ‘if  $A$  then  $B$ ’, we probably want *modus ponens* to be valid. So we’ll want the relevant Kripke models to be reflexive. Along the same lines, we could now look at other conditions on the accessibility relation and decide whether they should be imposed based on what they imply for the logic of conditionals. But let’s take a shortcut.

Above I suggested that sentence (1) can be understood to say that it is *impossible* to leave after 5 and still make it to the train. Impossible in what sense? There are many possible worlds at which we leave after 5 and still make it to the train; for example, worlds at which the train departs two hours later, or worlds at which we live right next to the station. When I say that it is impossible to leave after 5 and still make it to the train, I mean that it is impossible *given what we know about the departure time, our location, etc.*

Generalizing, a tempting proposal is that for indicative conditionals like (1), the accessibility relation is the epistemic accessibility relation we studied in chapter 5: a world is accessible from  $w$  iff it is compatible with what is known at  $w$ . The logic of

## 8. Conditionals

---

indicative conditionals is then determined by the logic of epistemic necessity; we don't need to figure out the relevant accessibility relations from scratch.

Since knowledge varies from agent to agent, the present idea implies that the truth-value of indicative conditionals should be agent-relative. This seems to be confirmed by the following puzzle, due to Allan Gibbard (1981).

Sly Pete and Mr. Stone are playing poker on a Mississippi riverboat. It is now up to Pete to call or fold. My henchman Zack sees Stone's hand, which is quite good, and signals its content to Pete. My henchman Jack sees both hands, and sees that Pete's hand is rather low, so that Stone's is the winning hand. At this point the room is cleared. A few minutes later, Zack slips me a note which says 'if Pete called, he won', and Jack slips me a note which says 'if Pete called, he lost'.

The puzzle is that Zack's note and Jack's note are intuitively contradictory, yet they both seem to be true.

We can resolve the puzzle if we understand the conditionals as strict conditionals with an agent-relative epistemic accessibility relation. Take Zack. Zack knows that Pete knows Stone's hand. Knowing that Pete would not call unless his hand is better, among the worlds compatible with Zack's knowledge, all worlds at which Pete calls are worlds at which Pete wins. So if  $p$  translates 'Pete called' and  $q$  'Pete won', then  $p \rightarrow q$  is true relative to Zack's information state. Relative to Jack's information state, however, the same sentence is false. Jack knows that Stone's hand is better than Pete's, but he doesn't know that Pete knows Stone's hand. So among the worlds compatible with Jack's knowledge, all worlds at which Pete calls are worlds at which Pete loses. Relative to Jack's information state,  $p \rightarrow \neg q$  is true.

Another advantage of the "epistemically strict" interpretation of indicative conditionals is that it explains why indicative conditionals with antecedents that are known to be false seem defective. For example, suppose Fred has gone to work. In that scenario, is (2) true or false?

- (2) If Fred has not gone to work, he is helping his neighbours.

The question is hard to answer, and not because we lack information about the scenario. Once we are told that Fred has gone to work, it is unclear how we are meant to assess whether Fred is helping his neighbours *if* he has not gone to work. On the epistemically strict interpretation, if  $A$  is known to be false, no  $A$ -world is

## 8. Conditionals

---

epistemically accessible, and so it is pointless to ask whether all accessible *A*-worlds are *B*-worlds.

So we have a promising alternative to the hypothesis that indicative conditionals are material conditionals. According to the alternative, they are epistemically strict conditionals: strict conditionals with an epistemic accessibility relation.

What about subjunctive conditionals? Return to the two Shakespeare conditionals from the previous section. When we evaluate the indicative sentence – ‘If Shakespeare didn’t write *Hamlet*, then someone else did’ – we hold fixed our knowledge that *Hamlet* exists; worlds where the play was never written are inaccessible. But when we evaluate the subjunctive sentence – ‘If Shakespeare hadn’t written *Hamlet*, then someone else would have’ – we consider worlds at which *Hamlet* was never written. If subjunctive conditionals are strict conditionals, this means that their accessibility relation does not track our knowledge or information. Unfortunately, as we are going to see in the next section, it is hard to say what else it could track.

This is one problem for the strict analysis of natural-language conditionals. Another problem lies in the logic of strict conditionals. Let’s have another look at the “paradoxes of material implication” from page 153. The strict analogues of **P1–P5** would go as follows:

- (P1S)  $B \models A \rightarrow B$
- (P2S)  $\neg A \models A \rightarrow B$
- (P3S)  $\neg(A \rightarrow B) \models A$
- (P4S)  $A \rightarrow B \models \neg B \rightarrow \neg A$
- (P5S)  $A \rightarrow B \models (A \wedge C) \rightarrow B$

Of these, **P1S–P3S** are easily seen to be false (on any plausible assumption about the accessibility relation). But **P4S** and **P5S** are true, no matter what we say about accessibility.

So if we want to faithfully formalize ordinary-language conditionals, we either have to explain away the apparent counterexamples to 4 and 5, or find a different translation.



**Exercise 8.4**

Show that **PS4** and **PS5** are true on the Kripke semantics for  $\rightarrow$  (for example, by giving a tree proof).

**Exercise 8.5**

In section 1, I gave examples showing that **P4** and **P5** sound wrong if indicative conditionals are translated as material conditionals (and so **P4S** and **P5S** sound wrong if indicative conditionals are translated as strict conditionals). Show that the problem also arises for subjunctive conditionals.

**Exercise 8.6**

A plausible norm of pragmatics is that a sentence should only be asserted if it is known to be true. If the logic of knowledge is at least S4, it follows that an epistemically strict conditional is assertable iff the corresponding material conditional is assertable. Explain.

### 8.3. Variably strict conditionals

I mentioned that the strict interpretation of conditionals has a problem with subjunctive conditionals. (In fact, the problem also arises for indicative conditionals, but it is easier to see with subjunctives.) The problem is best explained by an example.

As I am writing these notes, I am in Coombs Building, room 2228, with my desk facing the wall to Al Hájek's office in room 2229. In that context, (1) seems true.

- (1) If I were to drill a hole through the wall behind my desk, the hole would come out in Al's office.

There is no logical connection between the antecedent of (1) and the consequent. There are many possible worlds at which I drill a hole through the wall behind my desk and don't reach Al's office – for example, worlds at which my desk faces the opposite wall, worlds at which Al's office is in a different room, and so on. If we want to translate (1) as a strict conditional, all such worlds must be inaccessible.

Now consider (2).

## 8. Conditionals

---

(2) If the office spaces had been randomly reassigned yesterday, Al's office would (still) be next to mine.

(2) seems false, or at least extremely unlikely. But if (2) is a strict conditional, and worlds at which Al and I are in different offices are inaccessible – as they seem to be for (1) – then (2) should be true. Among the worlds at which I am in 2228 and Al is in 2229, all worlds at which the office spaces have been randomly reassigned yesterday are worlds at which Al's office is next to mine. When we evaluate (2), it looks like we no longer hold fixed who is in which office. Worlds that were inaccessible for (1) are accessible for (2).

So the accessibility relation for subjunctive conditionals appears to vary from conditional to conditional. As David Lewis put it, subjunctive conditionals seem to be not strict, but “variably strict”.

Let's try to get a better grip on how this might work. (What follows is a slightly simplified version of an analysis developed by Robert Stalnaker and David Lewis at around 1970.)

Intuitively, when we ask what would have been the case if a certain event had occurred, we are looking at worlds that are much like the actual world up to the time of the event, then deviate in some minimal way to allow the event to take place, and from then on unfold in accordance with the general laws of the actual world. For example, when we ask what would have happened if Shakespeare hadn't written *Hamlet*, we wonder what is the case at worlds that are much like the actual world until 1599, at which point some mundane circumstance prevents Shakespeare from writing *Hamlet*. Similarly for (1). Here we are considering worlds that are much like the actual world up to now, at which point I decide to drill a hole and find a suitable drill. These changes do not require my office to be in a different room, so worlds where I'm not in room 2228 are ignored. Figuratively speaking, such worlds are “too remote”: they differ from the actual world in ways that are not required to make the antecedent true.

On that picture, a subjunctive conditional is true iff the consequent is true at the *closest* worlds at which the antecedent is true – where closeness is a matter of similarity in certain respects. Let's assume that we have identified the relevant standards of similarity (a non-trivial task, but one we'll set aside). Let ' $v <_w u$ ' mean that  $v$  is closer to  $w$  than  $u$ , in the sense that  $v$  differs less than  $u$  from  $w$  in whatever respects are relevant to the interpretation of subjunctive conditionals.

We make the following structural assumptions about the world-relative ordering  $<$ .

## 8. Conditionals

---

1. If  $v <_w u$  then  $u \not<_w v$ . (Asymmetry)
2. If  $v <_w u$  and  $u <_w t$  then  $v <_w t$ . (Transitivity)
3. If  $v <_w u$ , then for all  $t$  either  $v <_w t$  or  $t <_w u$ . (Weak connectedness)
4. For all  $w$  and  $v$ ,  $v \not<_w w$ . (**Weak centring**)
5. For any non-empty set of worlds  $X$  and world  $w$  there is a  $v \in X$  such that there is no  $u \in X$  with  $u <_w v$ . (The **Limit Assumption**)

The first two assumptions are self-explanatory. The third is needed to ensure that the “equidistance” relation that holds between  $v$  and  $u$  if neither  $v <_w u$  nor  $u <_w v$  is an equivalence relation. With these three assumptions, we can picture each world  $w$  as surrounded by nested spheres of other worlds;  $v <_w u$  means that  $v$  is in a closer sphere around  $w$  than  $u$ .

Weak centring (assumption 4) means that every world is among the closest worlds to itself. Finally, the Limit Assumption ensures that for any consistent proposition  $A$  and world  $w$ , there is a set of closest  $A$ -worlds. Without the Limit Assumption, there could be an infinite chain of ever closer  $A$ -worlds, with no world being maximally close.

### Exercise 8.7

Define a weaker relation  $\leq$  so that  $v \leq_w u$  iff  $u \not<_w v$ . Informally,  $v \leq_w u$  means that  $v$  is at least as similar to  $w$  in the relevant respects as  $u$ . Can you express the above five conditions on  $<$  in terms of  $\leq$ ? (For example, Asymmetry turns into the assumption that if  $u \not<_w v$  then  $v \leq_w u$ .)

We now introduce a variably strict operator  $\Box \rightarrow$  so that  $A \Box \rightarrow B$  is true at a world iff  $B$  is true at the closest worlds at which  $A$  is true. Models for a language with the  $\Box \rightarrow$  operator must contain a closeness ordering  $<$  on the set of worlds.

### Definition 8.2

A **similarity model** consists of

- a non-empty set  $W$ ,
- for each  $w \in W$  an order  $<_w$  that satisfies the above five conditions, and
- a function  $V$  that assigns to each sentence letter and each member of  $W$  a truth-value.

## 8. Conditionals

---

For the semantics of  $\Box \rightarrow$ , we can re-use a concept of from section 6.3. Let  $S$  be an arbitrary set of worlds, and let  $w$  be some world (that may or may not lie in  $S$ ). Normally, some worlds in  $S$  will be more similar to  $w$  than others. It will be useful to have a term for the *most* similar worlds to  $w$ , among all worlds in  $S$ . That term is  $Min^{<w}(S)$ , which is officially defined as follows:

$$Min^{<w}(S) =_{\text{def}} \{v : v \in S \wedge \neg \exists u(u \in S \wedge u <_w v)\}.$$

Since  $\{u, M, u \models A\}$  is the set of worlds (in model  $M$ ) at which  $A$  is true,  $Min^{<w}(\{u : M, u \models A\})$  is the set of closest worlds to  $w$  at which  $A$  is true. We want  $A \Box \rightarrow B$  to be true at  $w$  iff  $B$  is true at the closest  $A$ -worlds to  $w$ . So:

**Definition 8.3: Similarity semantics for  $\Box \rightarrow$**

If  $M$  is a similarity model and  $w$  a world in  $M$ , then

$M, w \models A \Box \rightarrow B$  iff  $M, v \models B$  for all  $v \in Min^{<w}(\{u : M, u \models A\})$ .

To see this in action, let's verify that *modus ponens* is valid for  $\Box \rightarrow$ . That is, we want to show that if  $A \Box \rightarrow B$  and  $A$  are both true at some world  $w$  in some similarity model, then so is  $B$ . Since  $A \Box \rightarrow B$  is true at  $w$ , then by definition 8.3,  $B$  is true at all worlds in  $Min^{<w}(\{u : M, u \models A\})$ . We also know that  $A$  is true at  $w$ . By weak centring,  $w$  is one of the closest worlds to itself. So it is also one of the closest  $A$ -worlds to itself. That is,  $w$  is in  $Min^{<w}(\{u : M, u \models A\})$ . So  $B$  is true at  $w$ .

**Exercise 8.8**

Explain why  $A \Box \rightarrow B$  entails  $A \rightarrow B$ .

**Exercise 8.9**

Show that if  $A$  is true at no worlds, then  $A \Box \rightarrow B$  is true.

According to definition 8.3,  $A \Box \rightarrow B$  functions as a universal quantifier over certain worlds: the closest  $A$ -worlds. In some ways, the variably strict conditional therefore behaves like a box.

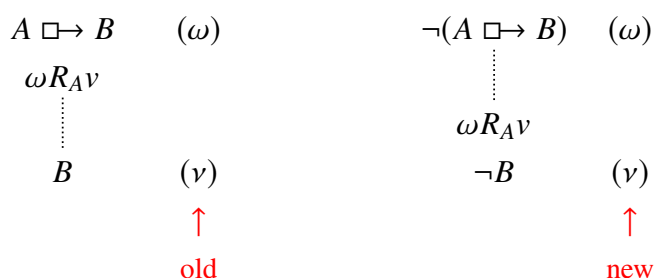
Let's say that for any sentence  $A$ , a world  $v$  is *A-accessible* from  $w$  (for short,

## 8. Conditionals

---

$wR_A v$ ) iff  $v$  is one of the closest  $A$ -worlds to  $w$ ; that is, iff  $v \in \text{Min}^{\prec w}(\{u : M, u \models A\})$ . Definition 8.3 then states that  $A \Box \rightarrow B$  is true at a world  $w$  iff  $B$  is true at all worlds  $A$ -accessible from  $w$ . These are the standard truth-conditions for the box, except that accessibility is relativised to the antecedent  $A$ .

We can therefore adapt the standard tree rules for the box to reason with  $\Box \rightarrow$ , as follows.



To get a complete tree system, we need further rules. For example, the above two rules don't account for the fact that the closest  $A$ -worlds are always  $A$ -worlds. They also don't account for the fact that every  $A$ -world is among the closest  $A$ -worlds to itself (by weak centring). We can add two more rules to fill these gaps.



The resulting tree system is still not complete, because it doesn't reflect interactions between different accessibility relations. For example, if the worlds that are  $A$ -accessible from some world  $w$  include  $B$ -worlds, then the  $A \wedge B$ -accessible worlds from  $w$  must be contained within the worlds  $A$ -accessible from  $w$ . The complete tree rules for the  $\Box \rightarrow$  operator turn out to be rather complicated. I will leave it at the above four rules, which suffice to establish many useful facts about variably strict conditionals.

Here is a tree proof to show (once more) that *modus ponens* is valid.

## 8. Conditionals

---

1.  $A \Box \rightarrow B$  (w) (Ass.)
  2.  $A$  (w) (Ass.)
  3.  $\neg B$  (w) (Ass.)
  4.  $wR_A w$  (Centring)
  5.  $B$  (w) (1,4)
- x

### Exercise 8.10

Give tree proofs for the following statements.

- (a)  $A, \neg B \models \neg(A \Box \rightarrow B)$
- (b)  $A \Box \rightarrow B, A \Box \rightarrow C \models A \Box \rightarrow (B \wedge C)$
- (c)  $A \Box \rightarrow (B \wedge C) \models (A \Box \rightarrow B) \wedge (A \Box \rightarrow C)$
- (d)  $A \Box \rightarrow \neg A \models A \Box \rightarrow B$

Since the tree rules I have presented are sound, you can be sure that whenever a tree closes then the tested entailment or validity holds. But since the rules are not complete, care is required when a tree doesn't close. You always need to check if a model read off from an open tree is an actual countermodel.

Constructing countermodels from open trees is at any rate not entirely straightforward. By way of illustration, let's show that  $A \Box \rightarrow B$  and  $B \Box \rightarrow C$  does not entail  $A \Box \rightarrow C$ . The tree starts like this.

1.  $A \Box \rightarrow B$  (w) (Ass.)
2.  $B \Box \rightarrow C$  (w) (Ass.)
3.  $\neg(A \Box \rightarrow C)$  (w) (Ass.) ✓
4.  $wR_A v$  (3)
5.  $\neg C$  (v) (3)
6.  $B$  (v) (1,4)
7.  $A$  (v) (4, Truth)

The Centring rule would allow us to add six more lines, but they wouldn't be useful, so let's stop here.

## 8. Conditionals

---

The open tree suggests that there is a countermodel with two worlds,  $w$  and  $v$ . At  $v$ ,  $A$  and  $B$  are true. We also have  $wR_A v$ , so  $v$  is among the closest  $A$ -worlds to  $w$ . Since  $wR_A w$  is not in the tree, let's assume that  $w$  is not among the closest  $A$ -worlds to  $w$ , which also means that  $A$  is false at  $w$ . We don't have  $wR_B v$  on the tree either. So even though  $B$  is true at  $v$ ,  $v$  is not among the closest  $B$ -worlds to  $w$ . We can ensure this by assuming that  $B$  is true at  $w$  itself, and that  $w$  is the unique closest  $B$ -world from itself. Now you can verify that  $A \Box \rightarrow B$  and  $B \Box \rightarrow C$  are both true at  $w$  while  $A \Box \rightarrow C$  is false.

The example also illustrates one of the many differences between  $\Box \rightarrow$  and  $\rightarrow$ : while  $A \rightarrow B$  and  $B \rightarrow C$  entail  $A \rightarrow C$ , the same inference with  $\Box \rightarrow$  is invalid. And arguably the inference *is* invalid with subjunctive conditionals. Stalnaker gives the following (cold-war era) counterexample.

If Hoover had been born a Russian, he would have been a Communist.  
 If Hoover were a Communist, he would have been be a traitor.  
 Therefore, If Hoover had been born a Russian, he would have been be a traitor.

### Exercise 8.11

Give counterexamples to the following, either by trying to construct them from open trees, or directly.

- (a)  $B \models A \Box \rightarrow B$
- (b)  $\neg A \models A \Box \rightarrow B$
- (c)  $\neg(A \Box \rightarrow B) \models A$
- (d)  $A \Box \rightarrow B \models \neg B \Box \rightarrow \neg A$
- (e)  $A \Box \rightarrow B \models (A \wedge C) \Box \rightarrow B$

As the preceding exercise shows, none of the “paradoxes of material implication” carry over to variably strict conditionals. In this respect,  $\Box \rightarrow$  seems to better match the ordinary use of conditionals than  $\rightarrow$ . On the other hand, you might have thought that the following entailment facts should hold, yet they do not. (The corresponding inferences with  $\rightarrow$  are valid; see exercise 8.2.)

1.  $((A \vee B) \Box \rightarrow C) \models (A \Box \rightarrow C) \wedge (B \Box \rightarrow C)$
2.  $A \Box \rightarrow (B \Box \rightarrow C) \models (A \wedge B) \Box \rightarrow C$

The semantics I have presented for  $\Box \rightarrow$  is a middle ground between those of Lewis and Stalnaker. Stalnaker assumes that  $<_w$  is not just weakly, but strongly connected: for any  $w, v, u$ , either  $v <_w u$  or  $v = u$  or  $u <_w v$ . This rules out ties in similarity and renders “Conditional Excluded Middle” valid:

$$\text{(CEM)} \quad (A \Box \rightarrow B) \vee (A \Box \rightarrow \neg B)$$

There is an ongoing controversy over whether conditionals in natural language validate **CEM**. On the one hand, it is natural to think that ‘it is not the case that if  $A$  then  $B$ ’ entails ‘if  $A$  then not  $B$ ’. On the other hand, suppose I have a number of coins in my pocket, none of which I have tossed. What would have happened if I had tossed one of the coins? Arguably, I might have gotten heads and I might have gotten tails. Either result is possible, but neither *would* have come about.

### Exercise 8.12

Show that the following statements are true on Stalnaker’s semantics:

- (a)  $A \wedge B \models A \Box \rightarrow B$
- (b)  $A \Box \rightarrow (B \vee C) \models (A \Box \rightarrow B) \vee (A \Box \rightarrow C)$

Lewis not only rejects strong connectedness, but also the Limit Assumption, arguing that there might well be an infinite chain of ever closer  $A$ -worlds. Definition 8.3 implies that if there are no closest  $A$ -worlds then any sentence of the form  $A \Box \rightarrow B$  is true. That does not seem right. Lewis therefore gives a more complicated semantics:

$M, w \models A \Box \rightarrow B$  iff either there is no  $v$  for which  $M, v \models A$  or there is some world  $v$  such that  $M, v \models A$  and for all  $u <_w v$ ,  $M, w \models A \rightarrow B$ .

It turns out that it makes no difference to the logic whether we impose the Limit Assumption and use the old definition or don’t impose the Limit Assumption and use Lewis’s new definition. The same sentences are valid either way.

## 8.4. The restrictor analysis

In section 6.3, we looked at conditional obligation statements like (1).



(1) If Jones is going to help his neighbours then he ought to tell them he's coming.

I claimed that this is best formalized not as  $p \rightarrow Oq$  or  $O(p \rightarrow q)$ , but as  $O(q/p)$ , where  $O(\cdot/\cdot)$  is a primitive two-place operator for conditional obligation. However, the original sentence (1) appears to contain a conditional ('if . . . then . . .') and the ordinary one-place operator 'ought'. One would like to know how these components work together to determine the intuitive meaning of (1).

To this end, we might reconsider if (1) can't be formalized with the help of a strict or variably strict conditional, say, as  $O(p \Box \rightarrow q)$ . Here I will explore the opposite strategy, of analysing *all* conditionals on the model of conditional obligation. This turns out to work so well that it has become the standard approach in linguistics.

Let's begin with a rather different case, to which David Lewis drew attention in 1975. Consider (2) and (3).

(2) If it rains, we always stay inside.

(3) If it rains, we sometimes stay inside.

Let *A* mean 'always' and *S* 'sometimes'. How could we translate (2) and (3)?

The "narrow scope" translations  $r \rightarrow A s$  and  $r \rightarrow S s$  are easily seen to be inadequate. For (2), the "wide scope"  $A(r \rightarrow s)$  looks good. One might expect that (3) should then be translated as  $S(r \rightarrow s)$ . But that is clearly wrong. (Note that  $S(r \rightarrow s)$  is equivalent to  $S(\neg r \vee s)$ .) Intuitively, (3) means that there are times at which it rains *and* we stay inside. So it should be translated as  $S(r \wedge s)$ . This is a bit surprising, because (3) seems to contain a conditional, yet in the translation we have a conjunction.

Things get worse if we turn to (4).

(4) If it rains, we mostly stay inside.

This says that among the occasions on which it rains, most are occasions on which we stay inside. Let *M* translate 'Mostly'. Neither  $M(r \rightarrow s)$  nor  $M(r \wedge s)$  capture the truth-conditions of (4), nor do  $M(r \Box \rightarrow s)$  or  $M(r \neg \rightarrow s)$ . Indeed, no statement of the form  $M A$  correctly translates (4).

We can formalize (4) if we treat *M* as a binary operator, taking two propositions as arguments:  $M(s/r)$ . The semantics for the two forms of *M* would look as follows, assuming that we are quantifying over times.

## 8. Conditionals

---

$$\begin{aligned}
 M, t \models MA & \quad \text{iff } M, s \models A \text{ for most times } s. \\
 M, t \models M(A/B) & \quad \text{iff } M, s \models A \text{ for most times } s \text{ such that } M, s \models B.
 \end{aligned}$$

So  $MA$  says that most times are  $A$ -times, while  $M(A/B)$  says that *among  $B$ -times*, most times are  $A$ -times. The function of the second argument to  $M(\cdot/\cdot)$ , which translates the ‘if’-clause in (4), is to **restrict the domain** of times over which the operator  $M$  quantifies.

If we assume that the ‘if’-clause in (4) is a restrictor of ‘mostly’, it is natural to assume that the ‘if’-clause in (2) and (3) plays the same role. (3) states that *among times when it rains*, we sometimes stay inside, while (2) states that among the same times, we always stay inside.

### Exercise 8.13

Translate ‘all dogs are barking’, ‘some dogs are barking’, into the language of predicate logic. Can you translate ‘most dogs are barking’ if you add a ‘most’ quantifier  $M$  so that  $MxA$  is true iff most things satisfy  $A$ ?

Now compare the semantics for  $O$  and  $O(\cdot/\cdot)$  from section 6.3.

$$\begin{aligned}
 M, w \models OA & \quad \text{iff } M, v \models A \text{ for all } v \in \text{Min}^{<w}(\{u : wRv\}) \\
 M, w \models O(B/A) & \quad \text{iff } M, v \models B \text{ for all } v \in \text{Min}^{<w}(\{u : wRv \text{ and } M, v \models A\})
 \end{aligned}$$

Informally,  $OA$  means that among the circumstantially accessible world, all the best worlds are  $A$ -worlds, while  $O(A/B)$  means that among the circumstantially accessible worlds *that are also  $B$ -worlds*, all the best worlds are  $A$ -worlds. The second argument place of  $O(\cdot/\cdot)$  restricts the domain of worlds over which  $O$  quantifies. As (2)–(4) illustrate, there is good reason to think that ‘if’-clauses in English can play this role.

The upshot is that when a sentence appears to contain a conditional and a modal operator, like (1) or (3) or (4), the ‘if’-clause sometimes serves to restrict the operator’s domain.

Angelika Kratzer forcefully argued that this is what if-clauses *always* do:

The history of the conditional is the story of a syntactic mistake. There is no two-place *if...then* connective in the logical forms of natural languages. *If*-clauses are devices for restricting the domains of various

operators. Whenever there is no explicit operator, we have to posit one.  
[Kratzer 1991]

Recall the following example from section 8.2.

(5) If we leave after 5, we will miss the train.

Here there doesn't seem to be any modal operator. According to Kratzer, we therefore have to posit a hidden operator. Perhaps the sentence contains a hidden epistemic necessity operator akin to 'must'. The if-clause restricts the domain of that operator. This makes (5) true iff *among the epistemically accessible worlds at which we leave after 5*, all worlds are worlds at which we miss the train. Equivalently: among the epistemically accessible worlds, all worlds are either worlds at which we don't leave after 5 or they are worlds at which we miss the train. This form of the restrictor analysis therefore yields the same truth-conditions for (5) as the strict epistemic analysis.

The case of subjunctive conditionals is similar. Here we might suggest that the 'if'-clause restricts a necessity operator with a non-epistemic, circumstantial flavour. The resulting truth-conditions might then be equivalent to those of the corresponding strict conditionals  $\Box(A \rightarrow B)$ .

But the equivalence depends on the interpretation of the necessity operator. Suppose we postulate a hidden operator  $\Box$  so that  $\Box A$  is true iff  $A$  is true at the *closest* of the accessible worlds (relative to some suitable ordering). This is how we interpreted  $O$  in section 6.3, reading "closeness" as betterness. Restricting the domain of this  $\Box$  operator effectively leads to the truth-conditions for variably strict conditionals.

The upshot is that the restrictor analysis does not determine a new logic of conditionals. Logically, 'if  $A$  then  $B$ ' may well behave just like  $A \rightarrow B$  or  $A \Box \rightarrow B$  (or even  $A \rightarrow B$ ). Nonetheless, the analysis is worth knowing, because it promises to shed light on many otherwise puzzling facts about conditionals in natural language. For example, it explains why 'if  $A$  then  $B$ ' is sometimes best translated as ' $A \wedge B$ ' (as in example (3)); it explains why 'if  $A$  then it must be that  $B$ ' can often be translated as  $\Box(A \rightarrow B)$ , even though 'must' appears to be in the consequent of the English sentence; and it explains why sometimes, what appear to be combinations of modal operators and conditionals are best translated with a primitive two-place operator (as in example (4)).