8 Conditionals

8.1 Material conditionals

We are often interested not just in whether something is in fact the case, but also in whether it is (or would be) the case *if* something else is (or would be) the case. We might, for example, wonder in what will happen to the climate if we don't reduce greenhouse gases, or whether World War 2 could have been avoided if certain steps had been taken in the 1930s.

A sentence stating that something is (or would be) the case if something else is (or would be) the case is called a **conditional**. What exactly, do these statements mean? What is their logic? Philosophers have puzzled over these questions for more than 2000 years, with no agreement in sight.

One attractively simple view is that a conditional 'if A then B' is true iff the antecedent A is false or the consequent B is true. This would make 'if A then B' equivalent to 'not A or B'. Conditionals with these truth-conditions are called **material conditionals**.

The "conditionals" $A \to B$ of classical logic are material. $A \to B$ is equivalent to $\neg A \lor B$. The "attractively simple" view that English conditionals are material conditionals would mean that we can faithfully translate English conditionals into $\mathfrak{L}_M$-sentences of the form $A \to B$. Is this correct?

There are some arguments for a positive answer. Suppose I make the following promise.

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

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, you could not fault me for breaking 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 true iff A is false or B is true.

Another argument for analysing English conditionals as material conditionals 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* inference.) Suppose I tell you that Nadia is either in Rome or in Paris. Trusting me, you can infer that if she's not in Rome then she's in Paris. Now we can reason as follows.

If A is true and B is false, then the conditional 'if A then B' is clearly false. Suppose, alternatively, that A is false or B is true. Then 'not-A or B' is true. By or-to-if, we can infer that 'if A then B' is true as well. Thus 'if A then B' is true iff A is false or B is true.

Despite these arguments, most philosophers and linguists don't think that English conditionals are material conditionals. Consider these facts about logical consequence (in classical propositional logic).

(M1) $B \models A \to B$

(M2) $\neg A \models A \to B$

(M3) $\neg(A \to B) \models A$

(M4) $A \to B \models \neg B \to \neg A$

(M5) $A \to B \models (A \land C) \to B$

If English conditionals were material conditionals then the following inferences, corresponding to (M1)–(M5), would be valid.

(E1) There won't be a nuclear war. Therefore: If Russia attacks the US with nuclear weapons then there won't be a nuclear war.

(E2) There won't be a nuclear war. Therefore: If there will be a nuclear war then nobody will die.

(E3) It is not the case that if it will rain tomorrow then the Moon will fall onto the Earth. Therefore: It will rain tomorrow.

(E4) 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.

(E5) 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.

These inferences do not sound good. If we wanted to defend the view that English conditionals are material conditionals we would have to explain why they sound bad even though they are valid. We will not explore this option any further.

Exercise 8.1

Can you find a different analysis of English conditionals that, like the material analysis, would make conditionals truth-functional, but that would render all of (E1)–(E5) invalid?

Even those who defend the material analysis of English conditionals admit that it does not work for all English conditionals. Consider (2).

(2) If water is heated to 100° 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° C, it evaporates. It is a quantified, or modal claim.

Another important class of conditionals that can't be analysed as material conditionals 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**. Intuitively, an indicative conditional states that something is *in fact* the case on the assumption that something else is the case. A subjunctive conditional like (4) states that something *would be* the case if something else *were* the case. Typically we know that the "something else" is not in fact the case. We know, for example, that Shakespeare wrote *Hamlet* and therefore that the antecedent of (4) is false. For this reason, subjunctive conditionals are also called *counterfactual conditionals* or simply *counterfactuals*.

It should be clear that subjunctive conditionals are not material conditionals. I said that (4) is almost certainly false. But it almost certainly has a false antecedent. So the corresponding material conditional is almost certainly true.

8.2 Strict conditionals

One apparent difference between material conditionals $A \to B$ and conditionals in natural language is that $A \to B$ 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, someone who utters (1) wants 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 $p \to q$ but as $\Box(p \to q)$ or, equivalently, $\neg\Diamond(p \land \neg q)$.

Sentences that are equivalent to $\Box(A \to B)$ are called **strict conditionals**. The label goes back to C.I. Lewis (1918), who also introduced the abbreviation $A \strictif B$ for $\Box(A \to B)$.

Lewis was not interested in 'if …then …' sentences. He introduced $A \strictif B$ to formalize 'A implies B' or 'A entails B'. His intended use of $\strictif$ 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 $\strictif$ is an *object-language* operator that, like $\land$ or $\to$, can be placed between any two sentences in a formal language to generate another sentence in the language. $p \strictif (q \strictif p)$ 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 $p \strictif q$ is true on some interpretation of the sentence letters and false on others. If p means that it raining heavily and q that it is raining, then $p \strictif q$ is true because the hypothesis that it is raining heavily implies that it is raining.

Let's set aside Lewis's project of formalizing the concept of implication. Our goal is to find an object-language construction that functions like 'if …then …' in English. To see whether '… $\strictif$ …' can do the job, let's have a closer look at the logic of strict conditionals.

Since $A \strictif B$ is equivalent to $\Box(A \to B)$, standard Kripke semantics for the box also provides a semantics for strict conditionals. In Kripke semantics, $\Box(A \to B)$ is true at a world w iff $A \to B$ is true at all worlds v accessible from w. And $A \to B$ is true at v iff A is false at v or B is true at v. We therefore have the following truth-conditions for strict conditionals.

Definition 8.1: Kripke semantics for $\strictif$

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

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

Exercise 8.2

$A \strictif B$ is equivalent to $\Box(A \to B)$. Can you find a sentence schema with $\strictif$ 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 put on the accessibility relation. Without any constraints, $\strictif$ does not validate *modus ponens*, in the sense that $A \strictif B$ and A together do not entail B. We can see this by translating $A \strictif B$ back into $\Box(A \to B)$ 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.	$\Box(A \to B)$	(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 more we can do.

Exercise 8.3

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

If we assume that the accessibility relation is reflexive, the tree closes:

4.	wRw		(Ref.)
5.	$A \to B$	(w)	(1,4)

6.	$\neg A$	(w)	(5)	7.	B	(w)	(5)
	×				×		

It is not hard to show that *modus ponens* for $\strictif$ is valid on all and only the reflexive frames. Reflexivity is precisely what we need to render *modus ponens* valid. And we probably want *modus ponens* to be valid for English conditionals. If A is true and B false, then the conditional 'if A then B' seems clearly false. So we'll want the relevant Kripke models to be reflexive.

Exercise 8.4

Using the tree method, and translating $A \strictif B$ into $\Box(A \to B)$, confirm that following claims hold, for all A, B, C.

(a) $\models_K A \strictif A$
(b) $A \strictif B \models_K \neg B \strictif \neg A$
(c) $A \strictif B \models_K (A \land C) \strictif B$
(d) $A \strictif B, B \strictif C \models_K A \strictif C$
(e) $(A \lor B) \strictif C \models_K (A \strictif C) \land (B \strictif C)$
(f) $A \strictif (B \strictif C) \models_T (A \land B) \strictif C$
(g) $A \strictif B \models_{S4} C \strictif (A \strictif B)$
(h) $((A \strictif B) \strictif C) \strictif (A \strictif B) \models_{S5} A \strictif B$

Which of these do you think are plausible if we assume that $A \strictif B$ translates indicative conditionals 'if A then B'?

We could now look at other conditions on the accessibility relation and decide whether they should be imposed, based on what they would imply for the logic of conditionals. But let's take a shortcut.

I have suggested that sentence (1) might be understood as saying 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. There are, for example, worlds at which the train departs two hours later, worlds at which we live right next to the station, and so on. When I say that it is impossible to leave after 5 and still make it to the train, I arguably mean that it is impossible *given what we know about the departure time, our location, etc.*

Generalizing, a tempting proposal is that the accessibility relation that is relevant for indicative conditionals like (1) is the epistemic accessibility relation that we studied in chapter **??**, where a world v is accessible from w iff it is compatible with what is known at w. On that hypothesis, the logic of indicative conditionals is determined by the logic of epistemic necessity. We don't need to figure out the relevant accessibility relation 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.

> 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. He also knows that Pete would not call unless he has the better hand. So among the worlds compatible with Zack's knowledge, all worlds at which Pete calls are worlds at which Pete wins. If p translates 'Pete called' and q 'Pete won', then $p \strictif q$ 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. Among the worlds compatible with Jack's knowledge, all worlds at which Pete calls are therefore worlds at which Pete loses. Relative to Jack's information state, $p \strictif \neg q$ is true.

Another advantage of the "epistemically strict" interpretation is that it might explain why indicative conditionals with antecedents that are known to be false seem defective. For example, imagine a scenario in which Jones has gone to work. In that scenario, is (2) true or false?

(2) If Jones has not gone to work then 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 Jones has gone to work, it is unclear how we are meant to assess whether Jones is helping his neighbours *if* he has not gone to work. On the epistemically strict interpretation, (2) says that Jones is helping his neighbours at all epistemically accessible worlds at which Jones hasn't gone to work. Since we know that Jones has gone to work, there are no epistemically accessible worlds at which he hasn't gone to work. And if there are no A-worlds then we naturally balk at the question whether all A-worlds are B-worlds. (In logic, we resolve to treat 'all As are B' as true if there are no As. Accordingly, (2) comes out true on the epistemically strict analysis. But we can still explain why it seems defective.)

We have found a promising alternative to the hypothesis that indicative conditionals are material conditionals. According to the present 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. That's why the conditional is true. At all accessible worlds at which Shakespeare didn't write *Hamlet*, someone else wrote the play. When we evaluate the subjunctive conditional – 'If Shakespeare hadn't written *Hamlet*, then someone else would have' – we do consider worlds at which *Hamlet* was never written, even though we know that the actual world is not of that kind. If subjunctive conditionals are strict conditionals, then 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. Remember (E1)–(E5) from page 44. If English conditionals are strict conditionals, then (E1)–(E3) are invalid. For example, while q entails $p \rightarrow q$, it does not entail $p \dashv\!\!\!3\; q$. But the strict analogues of (M4) and (M5) still hold, no matter what we say about accessibility (see exercise

3.2):

$$A \strictif B \models \neg B \strictif \neg A;$$
$$A \strictif B \models (A \land C) \strictif B.$$

So we still predict that the inferences (E4) and (E5) are valid.

(E4) 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.

(E5) 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.

Exercise 8.5

The badness of (E4) and (E5) suggests that indicative conditionals can't be analysed as strict conditionals. Can you give a similar argument suggesting that *subjunctive* conditionals can't be analysed as strict conditionals?

Exercise 8.6

A plausible norm of pragmatics is that a sentence should only be asserted if it is known to be true. Let's call a sentence *assertable* if it is known to be true. Show that if the logic of knowledge is at least S4, then an epistemically strict conditional $A \strictif B$ is assertable iff the corresponding material conditional $A \to B$ is assertable.

Exercise 8.7

Explain why the 'or-to-if' inference from 'p or q' to 'if not p then q' is invalid on the assumption that the conditional is epistemically strict. How could a friend of this assumption explain why the inference nonetheless looks reasonable, at least in normal situations? (Hint: Remember the previous exercise.)

8.3 Variably strict conditionals

Let's have a closer look at subjunctive conditionals. As I am writing these notes, I am sitting in Coombs Building, room 2228, with my desk facing the wall to Al Hájek's office in room 2229. In light of these facts, (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 (1) is a strict conditional then all such worlds must be inaccessible.

Now consider (2).

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

(2) seems false, or at least very unlikely. But if (2) is a strict conditional, and worlds at which Al is not in room 2229 or I am not in 2228 are inaccessible – as they seem to be for (1) – then (2) should be true. Among 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, at least 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 in the 1960s.)

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 these worlds deviate in some minimal way to allow the event to take place. Afterwards the worlds unfold in accordance with the general laws of the actual world.

For example, if we wonder what would have happened if Shakespeare hadn't written *Hamlet*, we are interested in worlds that are like the actual world until 1599, at which point some mundane circumstances prevent Shakespeare from writing *Hamlet*. We are not interested in worlds at which Shakespeare was never born, or in which the laws of nature are radically different from the laws at our world. One might reasonably judge that Shakespeare would have been a famous author even if he hadn't written *Hamlet*, although we would hardly be famous in worlds in which he was never born.

Likewise 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. Worlds where I'm not in room 2228 can be 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.

This suggests that 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. The closest worlds (to the actual world) at which Shakespeare didn't write *Hamlet* are worlds that almost perfectly match the actual world until 1599, then deviate a little so that Shakespeare didn't write Hamlet, and afterwards still resemble the actual world with respect to the general laws of nature. We will not try to spell out in full generality what the relevant closeness measure should look like.

Let '$v \prec_w u$' 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 $\prec$.

1. If $v \prec_w u$ then $u \nprec_w v$. (Asymmetry)
2. If $v \prec_w u$, then for all t either $v \prec_w t$ or $t \prec_w u$. (Quasi-connectedness)
3. 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 $u \prec_w v$.

Asymmetry is self-explanatory. Quasi-connectedness (a.k.a. negative transitivity) ensures that the "equidistance" relation that holds between v and u if neither $v \prec_w u$

nor $u \prec_w v$ is an equivalence relation. With these two assumptions, we can picture each world w as associated with nested spheres of worlds; $v \prec_w u$ means that v is in a more narrow w-sphere than u.

Assumption 3 is known as the **Limit Assumption**. It 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 8.8

Show that asymmetry and quasi-connectedness imply transitivity.

Exercise 8.9

Define $\preceq_w$ so that $v \preceq_w u$ iff $u \nprec_w v$ (that is, if it is not the case that $u \prec_w v$). Informally, $v \preceq_w u$ means that v is at least as similar to w in the relevant respects as u. Many authors use $\preceq$ rather than $\prec$ as their basic notion. Can you express the above three conditions on $\prec$ in terms of $\preceq$? (For example, Asymmetry turns into the assumption that for all w, v, u, either $u \preceq_w v$ or $v \preceq_w u$.)

We are going introduce a variably strict operator $\Box\!\!\rightarrow$ so that $A \Box\!\!\rightarrow B$ is true at a world w iff B is true at the closest worlds to w at which A is true. Models for a language with the $\Box\!\!\rightarrow$ operator must contain closeness orderings $\prec$ on the set of worlds.

Definition 8.2

A **similarity model** consists of

- a non-empty set W,
- for each w in W an asymmetric and quasi-connected order $\prec_w$ that satisfies the Limit Assumption, and
- a function V that assigns to each sentence letter a subset of W.

To formally state the semantics of $\Box\!\!\rightarrow$, we can re-use a concept from section 1.3. Let S be an arbitrary set of worlds, and let w be some world (that may or may nor be in S). It will be useful to have an expression that picks out the most similar worlds to w, among all the worlds in S. This expression is $Min^{<_w}(S)$, which we have defined as follows in section 1.3:

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

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

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

If M is a similarity model and w a world in M, then
$M, w \models A \mathrel{\Box\!\!\rightarrow} B$ iff $M, v \models B$ for all v in $Min^{<_w}(\{u : M, u \models A\})$.

You may notice that $A \mathrel{\Box\!\!\rightarrow} B$ works almost exactly like $\mathsf{O}(B/A)$ from section 1.3. There, I said that for any world w in any deontic ordering model M,

$$M, w \models \mathsf{O}(B/A) \text{ iff } M, v \models B \text{ for all } v \text{ in } Min^{<_w}(\{u : wRu \text{ and } M, u \models A\}).$$

The main difference is that conditional obligation is sensitive to an accessibility relation. If that relation is an equivalence relation then this makes no difference to the logic.

Of course, the order $<$ in deontic ordering models is supposed to represent degree of conformity to norms, while the order $<$ in similarity models represents a certain similarity ranking in the evaluation of subjunctive conditionals. A different type of ordering might be in play when we evaluate indicative conditionals, which some have argued should also be interpreted as variably strict. But again, these differences in interpretation don't affect the logic.

Suppose we add the $\Box\!\!\rightarrow$ operator to the language of standard propositional logic. The set of sentences in this language that are true at all worlds in all similarity models is known as **system V**. There are tree rules and axiomatic calculi for this system, but they aren't very user-friendly. We will only explore the system semantically.

To begin, we can check whether *modus ponens* is valid for $\Box\!\!\rightarrow$. That is, we check whether the truth of A and $A \mathrel{\Box\!\!\rightarrow} B$ at a world in a similarity model entails the truth of B.

Assume that A and $A \mathrel{\Box\!\!\rightarrow} B$ are true at a world w. By definition 3.3, the latter means that B is true at all the closest A-worlds to w (at all worlds in $Min^{\prec_w}(\{u : M, u \models A\})$). The world w itself is an A-world. If we could show that w is among the closest A-worlds to itself then we could infer that A is true at w.

Without further assumptions, however, we can't show this. If we want to validate *modus ponens*, we must add a further constraint on our models: that every world is among the closest worlds to itself. More precisely,

$$\text{for all worlds } w \text{ and } v,\ v \nprec_w w.$$

This assumption is known as **Weak Centring**. The logic we get if we impose this constraint is **system VC**.

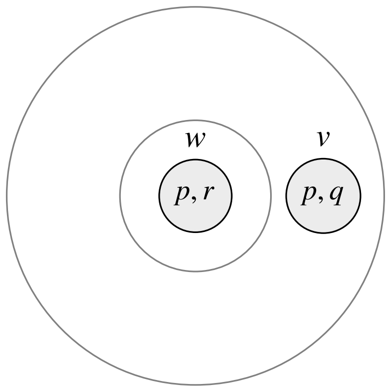
Exercise 8.10

Should we accept Weak Centring for deontic ordering models?

Exercise 8.11

Explain why $A \mathrel{\Box\!\!\rightarrow} B$ entails $A \rightarrow B$, assuming Weak Centring.

None of the problematic inferences (E1)–(E5) are valid if the relevant conditionals are interpreted as variably strict. (E5), for example, would assume that $p \mathrel{\Box\!\!\rightarrow} r$ entails $(p \land q) \mathrel{\Box\!\!\rightarrow} r$. But it does not. We can give a countermodel with two worlds w and v; p is true at both worlds, q is true only at v, and r only at w; if w is closer to itself than v, then $p \mathrel{\Box\!\!\rightarrow} r$ is true at w (because the closest p-worlds to w are all r-worlds), but $(p \land q) \mathrel{\Box\!\!\rightarrow} r$ is false at w (because the closest $(p \land q)$-worlds to w aren't all r-worlds).



The diagram on the right represents this model. The circles around w depict the similarity spheres. w is closer to w than v because it is in the innermost sphere around w, while v is only in the second sphere. (If v were also in the innermost sphere then the two worlds would be equally close to w. That's allowed.) In general, we can represent the assumption that a world v is closer to a world w than a world

u ($v \prec_w u$) by putting v is in a closer sphere around w than u. I have not drawn any spheres around v because it doesn't matter what these look like.

Exercise 8.12

Draw countermodels showing that (E1)–(E4) are invalid if the conditionals are translated as statements of the form $A \mathrel{\Box\!\!\rightarrow} B$. (Hint: You never need more than two worlds.)

The logic of variably strict conditionals is weaker than the logic of strict conditionals. Some have argued that it is too weak to explain our reasoning with conditionals. It is, for example, not hard to see that the following statements are all false. (The corresponding statements for $\dashv$ are true; see exercise 3.2.)

1. $p \mathrel{\Box\!\!\rightarrow} q, q \mathrel{\Box\!\!\rightarrow} r \models p \mathrel{\Box\!\!\rightarrow} r$
2. $((p \lor q) \mathrel{\Box\!\!\rightarrow} r) \models (p \mathrel{\Box\!\!\rightarrow} r) \land (q \mathrel{\Box\!\!\rightarrow} r)$
3. $p \mathrel{\Box\!\!\rightarrow} (q \mathrel{\Box\!\!\rightarrow} r) \models (p \land q) \mathrel{\Box\!\!\rightarrow} r$

If English conditionals are variably strict, this means (for example) that we can't infer 'if p then r' from 'if p then q' and 'if q then r'. But isn't this a valid inference?

Well, perhaps not. Stalnaker gave the following counterexample, using cold-war era subjunctive conditionals.

> If J. Edgar Hoover had been born a Russian, he would be a communist.
> If Hoover were a communist, he would be a traitor.
> Therefore, if Hoover had been born a Russian, he would be a traitor.

Exercise 8.13

Can you find a case where 'if p or q then r' does not appear to entail 'if p then r' and 'if q then r'? You can use either indicative or subjunctive conditionals. (Hint: Try to find a case in which 'if p or q then p' sounds acceptable.)

The semantics I have presented for $\mathrel{\Box\!\!\rightarrow}$ is a middle ground between that of Lewis and Stalnaker. Stalnaker assumes that $\prec_w$ is not just quasi-connected, but connected: for any w, v, u, either $v \prec_w u$ or $v = u$ or $u \prec_w v$. ('$v = u$' means that v and u are the same world.) This rules out ties in similarity: no sphere contains more than one world.

Stalnaker's logic (called **C2**) is stronger than Lewis's VC. The following principle of "Conditional Excluded Middle" is C2-valid but not VC-valid:

(CEM) $(A \mathrel{\Box\!\!\rightarrow} B) \lor (A \mathrel{\Box\!\!\rightarrow} \neg B)$

Whether conditionals in natural language satisfy Conditional Excluded Middle is a matter of ongoing debate. On the one hand, it is natural think that 'it is not the case that if p then q' entails 'if p then not q', which suggests that the principle is valid. 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 8.14

Explain why the following statements are true, for all A, B, C:

(a) $A \land B \models_{C2} A \mathrel{\Box\!\!\rightarrow} B$
(b) $A \mathrel{\Box\!\!\rightarrow} (B \lor C) \models_{C2} (A \mathrel{\Box\!\!\rightarrow} B) \lor (A \mathrel{\Box\!\!\rightarrow} C)$

Lewis not only rejects connectedness, but also the Limit Assumption. He argued that there might be an infinite chain of ever closer A-worlds. Definition 3.3 implies that if there are no closest A-worlds then any sentence of the form $A \mathrel{\Box\!\!\rightarrow} B$ is true. That does not seem right. Lewis therefore gives a more complicated semantics:

> $M, w \models A \mathrel{\Box\!\!\rightarrow} B$ 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 $u \prec_w v$, $M, w \models A \rightarrow B$.

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 Restrictors

Consider these two statements.

(1) If it rains we always stay inside.

(2) If it rains we sometimes stay inside.

On its most natural reading, (1) says that we stay inside at all times at which it rains. We can express this in $\mathfrak{L}_M$, using the box as a universal quantifier over the relevant times. (So $\Box A$ now means 'always A'.) The translation would be $\Box(r \to s)$.

One might expect that (2) should then be translated as $\Diamond(r \to s)$, where the diamond is an existential quantifier over the relevant times ('sometimes'). But $\Diamond(r \to s)$ is equivalent to $\Diamond(\neg r \vee s)$. This is true whenever $\Diamond\neg r$ is true. (2), however, isn't true simply because it doesn't always rain. On its most salient reading, (2) says there are times at which it rains *and* we stay inside. Its correct translation is $\Diamond(r \wedge s)$.

This is a little surprising, given that (2) seems to contain a conditional. Does the conditional here express a conjunction?

Things get worse if we look at (3).

(3) If it rains we usually stay inside.

Let's introduce an operator M for 'usually', so that $\mathsf{M}A$ is true at a time iff A is true at *most* times. Can you translate (3) with the help of M?

You can't. Neither $\mathsf{M}(r \to s)$ nor $\mathsf{M}(r \wedge s)$ capture the intended meaning of (3). $\mathsf{M}(r \wedge s)$ entails that r is usually true. But (3) doesn't entail that it usually rains. $\mathsf{M}(r \to s)$ is true as long as r is usually false, even if we're always outside when it is raining. You could try to bring in some of the new kinds of conditional that we've encountered in the previous sections. How about $\mathsf{M}(r \mathrel{\Box\!\!\to} s)$, or $\mathsf{M}(r \mathrel{-\!\!\!3} s)$, or $r \mathrel{\Box\!\!\to} \mathsf{M}s$, or $r \mathrel{-\!\!\!3} \mathsf{M}s$? None of these are adequate.

The problem is that (3) doesn't say, of any particular proposition, that it is true at most times. It doesn't say that among all times, most are such-and-such. Rather, it says that *among times at which it rains*, most times are times at which we stay inside. The function of the 'if'-clause in (3) is to **restrict the domain** of times over which the 'usually' operator quantifies.

Now return to (1) and (2). Suppose that here, too, the 'if'-clause serves to restrict the domain of times, so that 'always' and 'sometimes' only quantify over times at which it rains. On that hypothesis, (1) says that *among times at which it rains*, all times are times at which we stay inside, and (2) says that *among times at which it rains*, some times are times at which we stay inside. This is indeed what (1) and (2) mean, on their most salient interpretation.

As it turns out, 'among r-times, all times are s-times' is equivalent to 'all times are not-r-times or s-times'. That's why we can formalize (1) as $\Box(r \to s)$. 'Among

r-times, some times are s-times', on the other hand, is equivalent to 'some times are r-times and s-times'. That's why we can formalize (2) as $\Diamond(r \wedge s)$. It would be wrong to think that the conditional in (1) is material, the conditional in (2) is a conjunction, and the conditional in (3) is something else altogether. A much better explanation is that the 'if'-clause in (1) does the exact same thing as in (2) and (3). In each case, it restricts the domain of times over which the relevant operators quantify.

We can arguably see the same effect in (4) and (5).

(4) If the lights are on, Ada must be in her office.

(5) If the lights are on, Ada might be in her office.

Letting the box express epistemic necessity, we can translate (4) as $\Box(p \to q)$. But (5) can't be translated as $\Diamond(p \to q)$, which would be equivalent to $\Diamond(\neg p \vee q)$. Nor can we translate (5) as $p \to \Diamond q$, which is entailed by $\Diamond q$. It is easy to think of scenarios in which (5) is false even though 'Ada might be in her office' is true. The correct translation of (5) is plausibly $\Diamond(p \wedge q)$. The sentence is true iff there is an epistemically accessible world at which the lights are on and Ada is in her office.

As before, we can understand what is going if we assume that the 'if'-clause in (4) and (5) functions as a restrictor. The 'if'-clause restricts the domain of worlds over which 'must' and 'might' quantify. (4) says that *among epistemically possible worlds at which the lights are on*, all worlds are worlds at which Ada is in her office. (5) says that *among epistemically possible worlds at which the lights are on*, some worlds are worlds at which Ada is in her office.

Exercise 8.15

Translate 'all dogs are barking' and '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 $\mathsf{M}xFx$ is true iff most things satisfy Fx?

The hypothesis that 'if'-clauses are restrictors also sheds light on the problem of conditional obligation.

(6) Jones ought to help his neighbours.

(7) If Jones doesn't help his neighbours, he ought to not tell them that he's coming.

In chapter 1, we analyzed 'ought' as a quantifier over the best of the circumstantially accessible worlds. On this approach, (6) says that among the accessible worlds, all the best ones are worlds at which Jones helps his neighbours. Suppose the 'if'-clause in (7) serves to restrict the domain of worlds, excluding worlds at which Jones helps his neighbours. We then predict (7) to state that *among the accessible worlds at which Jones doesn't help his neighbours*, all the best worlds are worlds at which Jones doesn't tell his neighbours that he's coming. This can't be expressed by combining the monadic O quantifier with truth-functional connectives. Hence we had to introduce a primitive binary operator $\mathsf{O}(\cdot/\cdot)$.

The upshot of all this is that we can make sense of a wide range of puzzling phenomena by assuming that 'if'-clauses are restrictors. Their function is to restrict the domain or worlds or times over which modal operators quantify.

What, then, is the purpose of 'if'-clauses in "bare" conditionals like (8) and (9), where there are no modal operators to restrict?

(8) If Shakespeare didn't write *Hamlet*, then someone else did.

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

Here opinions vary. One possibility, prominently defended by the linguist Angelika Kratzer, is that even bare conditionals contain modal operators. Arguably, 'would' in (9) functions as a kind of box. If this box is a simple quantifier over circumstantially accessible worlds, and the 'if'-clause in (9) restricts its domain, then (9) can be formalized as $\Box(p \to q)$. If, on the other hand, 'would' in (9) works more light 'ought' – if it quantifyies over the *closest* of the accessible worlds –, and the 'if'-clause restricts the domain of accessible worlds, then the resulting truth-conditions are those of $p \mathrel{\Box\!\!\rightarrow} q$. Both the strict and the variably strict analysis of (9) are therefore compatible with the hypothesis that 'if'-clauses are restrictors.

What about (8)? This sentence really doesn't appear to contain a relevant modal. Kratzer suggests that it contains an unpronounced epistemic 'must': (8) says that if Shakespeare didn't write *Hamlet* then someone else *must* have written *Hamlet*. Assuming that the 'if'-clause restricts the domain of this operator, bare indicative conditionals would be equivalent to strict epistemic conditionals.

Exercise 8.16

Suppose bare indicative conditionals like (8) contain a box operator $\Box$ whose accessibility relation relates each world to itself and to no other world. (This is a redundant operator insofar as $\Box A$ is equivalent to A.) Assume the 'if'-clause restricts the domain of that operator. What are the resulting truth-conditions of (8)?

Exercise 8.17

Besides "would counterfactuals" there are also "might counterfactuals" like

(10) If I had played the lottery, I might have won.

Suppose 'might' is the dual of 'would', and suppose the 'if'-clause in (10) restricts the domain of worlds over which 'might' quantifies. It follows that 'if A then might B' is true iff B holds at some of the closest/accessible A-worlds. ('Closest' or 'accessible' depending on how we understand the 'would'/'might' operators.) Can you see why this casts doubt on the validity of Conditional Excluded Middle?