17 The truth-functional conditional

We have seen (at some length!) how to evaluate arguments whose essential logical materials are the three connectives 'and', 'or' and 'not'. We now ask: can the techniques that we have developed be smoothly extended to deal with arguments involving that other quite fundamental propositional connective, 'if'?

17.1 Some arguments involving conditionals

(a) Consider the following elementary arguments:

A If Jack bet on Eclipse, he lost his money. Jack did bet on Eclipse. So Jack lost his money.

B If Jack bet on Eclipse, he lost his money. Jack did not lose his money. So Jack did not bet on Eclipse.

C If Jack bet on Eclipse, he lost his money. So if Jack didn't lose his money, he didn't bet on Eclipse.

D If Jack bet on Eclipse, then he lost his money. If Jack lost his money, then he had to walk home. So if Jack bet on Eclipse, he had to walk home.

E Either Jack bet on Eclipse or on Pegasus. If Jack bet on Eclipse, he lost his money. If Jack bet on Pegasus, he lost his money. So, Jack lost his money.

Each of these first five arguments involving conditionals is intuitively valid. Contrast the next three arguments:

F If Jack bet on Eclipse, he lost his money. Jack lost his money. So Jack bet on Eclipse.

G If Jack bet on Eclipse, he lost his money. Jack did not bet on Eclipse. So Jack did not lose his money.

H If Jack bet on Eclipse, he lost his money. So if Jack did not bet on Eclipse, he did not lose his money.

These are plainly fallacious arguments. Suppose foolish Jack lost his money by betting not on Eclipse but on Pegasus, an equally hopeless horse. Then in each of these three cases, the premisses of the argument can be true and conclusion false.

The validity or invalidity of the arguments **A** to **H** has nothing specifically to do with Jack or betting, but is evidently due to the meaning of the conditional construction introduced by 'if'. We would therefore like to have a general way of handling arguments like these which rely on conditionals. Now, 'if' seems to be a binary connective like 'and' and 'or', at least in respect of combining two propositions to form a new one. So the obvious question is: can we carry over our methods for evaluating arguments involving those other binary connectives to deal with conditionals as well?

17.2 Four basic principles

As we will see in the next chapter, conditionals come in importantly different flavours; but in this chapter let's concentrate on simple cases like those in the previous section (leaving it open for now what counts as 'simple').

We will regard *if A, C* and *if A then C* as stylistic variants, treating the 'then' here as no more than helpful punctuation (perhaps the 'then' does more work in some conditionals; but put any such cases aside). And we now introduce some standard terminology:

Given a conditional *if A then C*, we refer to the 'if' clause *A* as the *antecedent* of the conditional, and to the other clause *C* as the *consequent*.

The *converse* of *if A then C* is the conditional *if C then A*.

The *contrapositive* of *if A then C* is the conditional *if not-C then not-A*.

We will also use the same terminology when talking about the formal analogues of vernacular conditionals.

Generalizing from the examples **A** and **B**, the following pair of basic principles of conditional reasoning evidently hold good (at least for simple conditionals):

(MP) An inference step of the form *A, if A then C, so C* is valid. This mode of inference is standardly referred to as *modus ponens*.

(MT) An inference step of the form *not-C, if A then C, so not-A* is valid. This mode of inference is standardly referred to as *modus tollens*.

And here is a closely related principle – we can call it the *falsehood condition*:

(FC) A conditional *if A then C* must be false if in fact *A* is true and *C* is false.

That's also obvious.

It is worth noting that both (FC) and (MT) are immediate consequences of (MP). Take three propositions of the form (i) *A*, (ii) *if A then C*, (iii) *not-C*. These are inconsistent since, by (MP), (i) and (ii) imply *C*, which contradicts (iii). Hence, if we keep (ii) and (iii) we have to reject (i), which is the principle (MT). And if we keep (i) and (iii) we have to reject (ii), which is the principle (FC).

Later, in §17.7, we will meet another core principle about conditionals: roughly, if we can infer *C* from the temporary assumption *A*, then we can drop the assumption and assert *if A then C*. But let's put that idea on hold for the moment, and note instead another basic fact about conditionals – namely, that they are not usually reversible:

(NR) From a conditional premiss of the form *if A then C*, we can't usually infer the converse conditional *if C then A*.

No doubt, if I have won a Nobel prize, then I am clever. But it doesn't follow that if I am clever, then I have won a Nobel prize.

17.3 Introducing the truth-functional conditional

(a) In headline terms, our Chapter 15 technique for evaluating arguments involving 'and', 'or' and 'not' has two stages, (1) translation into a PL language and (2) assessing the translated version for tautological validity by using a truth-table test.

Step (1) presupposes that the ordinary language connectives are close enough in their core meanings to the corresponding connectives of PL for logically important features of the original argument to get carried over by the translation.

Step (2) then depends on the fact that the PL connectives are truth-functional. That's needed to ensure that, given any valuation of the relevant propositional atoms, we can work out whether it makes the premisses of the PL argument true and conclusion false.

Hence, if our truth-table technique is to be straightforwardly extended to cover arguments involving the conditional, we need to be able to render conditional propositions into a formal language using a suitable truth-functional connective, and to do this in a way which preserves enough of the core meaning of the conditional.

(b) It is quite easy to see that there is only one truth-functional connective which satisfies inferential principles parallel to the four headlined principles (MP) to (NR) which apply to ordinary conditionals. Why so?

α	γ	$(\alpha \to \gamma)$
T	T	?
T	F	F
F	T	?
F	F	?

Suppose we add the symbol '$\to$' to a PL-style language in order to represent a connective which is intended to both truth-functional and conditional-like. The issue is how to fill in its truth-table. With an eye on (FC), we have already completed the second line: the supposed truth-functional conditional has to be false when it has a true antecedent and false consequent. So how does the rest of the table go? We'll give two arguments, slow and fast, for the same completion.

(c) Taking it slowly, step by step, note that the first line on the truth-table has to be completed with 'T'. Otherwise $(\alpha \to \gamma)$ would always be false when α is true. But of course there can be true conditionals with true antecedents – else we'd never be able to use (MP) with two true premisses $(\alpha \to \gamma)$ and α in order to prove that γ.

The last line also has to be completed with 'T'. Otherwise $(\alpha \to \gamma)$ would always be false when γ is false. But of course there can be true conditionals with false consequents – else we'd never be able to use (MT) with two true premisses $(\alpha \to \gamma)$ and $\neg\gamma$ to prove that $\neg\alpha$.

α	γ	$(\alpha \to \gamma)$
T	T	T
T	F	F
F	T	?
F	F	T

Which just leaves one entry for the truth-table to be decided. But if we put 'F' on the third line, then $(\alpha \to \gamma)$ would have the same truth-table as $(\gamma \to \alpha)$, making our supposed conditional reversible, contrary to the requirement that a conditional-like connective satisfies the analogue of (NR). Hence:

The only candidate for a conditional-like truth-function is the one defined by the following truth-table:

α	γ	$(\alpha \to \gamma)$
T	T	T
T	F	F
F	T	T
F	F	T

(d) For a fast track argument, take the wff '((P ∧ Q) → P)'. Assuming '→' is conditional-like, this should *always* be true (since necessarily, if 'P' and 'Q' are both true, then 'P' in particular is true). But the values of the antecedent/consequent in this wff can be any of T/T, or F/T, or F/F (depending on the values of 'P' and 'Q'). And we've just seen that the wff needs to evaluate as T in each case. So that forces the same completion of the table for our conditional-like truth-function.

(e) The ancient Stoic logician Philo (4th century BC) claimed, in effect, that the ordinary-language conditional has this truth-table. So the truth-functional connective here is sometimes still referred to as the *Philonian* conditional. It was reintroduced into modern logic by Frege in the nineteenth century, and then taken up by Bertrand Russell who called it – for bad reasons that need not detain us – *material implication*. The connective is now commonly known as the *material conditional* or simply as the *truth-functional conditional*.

17.4 Ways in which '→' is conditional-like

(a) Since PL languages can express all truth-functions, they can already express the truth-functional conditional. But it is conventional to add a special symbol for this truth-function, adjusting the syntactic and semantic rules to get richer languages we will call PLC languages. The way to do this is *exactly* as you would expect, as you will see in §17.6, so we do not need to pause over the details just yet. So let's simply continue to use the symbol '→' for this truth-function, assuming it is now officially part of the formal languages we are using.

How well does the material conditional behave as a formal counterpart to the ordinary-language conditional? We make a start on discussing this troublesome issue by confirming that if we transcribe the various examples in §17.1 using '→' for the conditional, and run truth-table tests, we do get the right verdicts. (Remember from §15.6 that we can extend the notion of tautological entailment and the use of truth-table tests beyond basic PL languages, so long as we are still dealing with truth-functional connectives.)

(b) Start with the modus ponens inference in

A If Jack bet on Eclipse, he lost his money. Jack did bet on Eclipse. So Jack lost his money.

Rendered into a PLC language with a suitable glossary, this straightfowardly goes into

A′ $(\mathsf{P} \to \mathsf{Q}),\ \mathsf{P}\ \therefore\ \mathsf{Q}$.

And running a full truth-table test (not exactly hard!), we get

P	Q	(P → Q)	P	Q
T	T	T	T	T
T	F	F	T	F
F	T	T	F	T
F	F	T	F	F

There are no bad lines with true premisses and false conclusion. Therefore argument **A′** is tautologically valid. Hence it is plain valid – as was the original version **A**.

We will skip the modus tollens argument **B** – it is a trivial exercise to confirm that the obvious formal rendition is tautologically valid too. So let's next look at

C If Jack bet on Eclipse, he lost his money. So if Jack didn't lose his money, he didn't bet on Eclipse.

This can be rendered as

C′ $(\mathsf{P} \to \mathsf{Q})\ \therefore\ (\neg\mathsf{Q} \to \neg\mathsf{P})$.

Running another truth-table test, this time we get

P	Q	(P → Q)	(¬Q → ¬P)
T	T	T	T
T	F	F	F
F	T	T	T
F	F	T	T

So again the inference in **C′** is tautologically valid and hence valid.

And there's more: that table shows that a material conditional not only tautologically entails its contrapositive, but is tautologically *equivalent* to it. But this little result still tracks the behaviour of ordinary-language conditionals, for the following is also true:

> Simple ordinary-language conditionals are equivalent to their contrapositives.

For example, 'If Jack bet on Eclipse, he lost his money' not only entails but is entailed by 'if Jack didn't lose his money, he didn't bet on Eclipse'.

Next example:

D If Jack bet on Eclipse, then Jack lost his money. If Jack lost his money, then Jack had to walk home. So if Jack bet on Eclipse, Jack had to walk home.

This goes into our PLC language as, say,

D′ $(\mathsf{P} \to \mathsf{Q}),\ (\mathsf{Q} \to \mathsf{R})\ \therefore\ (\mathsf{P} \to \mathsf{R})$.

Making a truth table using the now familiar shortcuts (evaluating the conclusion first, and then the premisses in turn but only as needed), we very quickly get

P	Q	R	(P → Q)	(Q → R)	(P → R)
T	T	T			T
T	T	F	T	F	F
T	F	T			T
T	F	F	F		F
F	T	T			T
F	T	F			T
F	F	T			T
F	F	F			T

There are no bad lines, so the argument is tautologically valid, and therefore plain valid, corresponding again to an intuitively valid vernacular argument.

The next example illustrates another common form of valid inference, a version of proof by cases:

E Either Jack bet on Eclipse or on Pegasus. If Jack bet on Eclipse, he lost his money. If Jack bet on Pegasus, he lost his money. So, Jack lost his money.

Using the material conditional, we can transcribe this as follows:

E′ $(P \lor Q),\ (P \to R),\ (Q \to R)\ \therefore\ R.$

You won't be surprised to learn that this too it tautologically valid. Here's a truth-table to confirm that (again, we evaluate the conclusion first, ignore good lines, and then look at the premisses in order, as needed).

P	Q	R	(P ∨ Q)	(P → R)	(Q → R)	R
T	T	T				T
T	T	F	T	F		F
T	F	T				T
T	F	F	T	F		F
F	T	T				T
F	T	F	T	T	F	F
F	F	T				T
F	F	F	F			F

(c) So much, then, for our first five example arguments **A** to **E**: they are both intuitively valid and valid by the truth-table test when rendered into suitable PLC languages. Let's turn, then, to the argument

F If Jack bet on Eclipse, he lost his money. Jack lost his money. So Jack bet on Eclipse.

As we noted before, the inference here of the form *If A then C, C, so A* is an instance of a gross fallacy (it's traditionally called *affirming the consequent*).

Take a translation of **F** using the material conditional, for instance

F′ $(P \to Q),\ Q\ \therefore\ P.$

This inference is tautologically invalid. Just consider the following trivial truth-table:

P	Q	(P → Q)	Q	P
T	T	T	T	T
T	F	F	F	T
F	T	T	T	F
F	F	T	F	F

And now note that the bad line here evidently corresponds to a possible state of affairs – i.e. Jack's not betting on Eclipse yet losing his money all the same. The possibility of that situation confirms that the original argument is plain invalid. So the tautological invalidity of **F′** directly reveals the invalidity of **F**.

Similarly, consider again:

G If Jack bet on Eclipse, he lost his money. Jack did not bet on Eclipse. So Jack did not lose his money.

This inference-step of the form *If A then C, not-A, so not-C* commits another horrible fallacy (traditionally called *denying the antecedent*). Transcribing into our formal language using the material conditional and running a truth-table test confirms this.

Finally note example **H**. Rendered into a PLC language, this goes into the tautologically invalid

H′ $(\mathsf{P} \to \mathsf{Q})\ \therefore\ (\neg\mathsf{P} \to \neg\mathsf{Q})$.

Be very, *very*, careful to distinguish the horribly fallacious inference **H′** from the perfectly valid inference from the same conditional to its contrapositive '$(\neg\mathsf{Q} \to \neg\mathsf{P})$' as in **C′**!

(d) Examples **A** to **H** show that rendering some simple arguments involving conditionals into a PLC language with its truth-functional conditional, and then using the truth-table test, can yield just the right verdicts about the validity or invalidity of the original arguments. And in the now familiar sort of way, we can generalize from those examples. So we see that – as intended – the material conditional satisfies the principles (MP), (MT), (FC) and (NR) from §17.2 with '→' now in place of 'if'.

In interim summary, then:

> The logical behaviour of '→' in PLC languages is indeed parallel to that of the vernacular 'if', in at least *some* cases, in *some* key respects.

17.5 'Only if' (and the biconditional)

(a) Up to now, we have been discussing conditionals of the form

(1) If *A* then *C*.

(with or without the 'then'). We can also write these 'if' conditionals in the form

(1′) *C*, if *A*.

But we will take this form, without further ado, to be a mere stylistic variant.

However, there is another basic kind of conditional which needs to be discussed, namely 'only if' conditionals. Consider the pair

(2) *A* only if *C*.
(2′) Only if *C*, *A*

Again, we can take these versions to be mere stylistic variants of each other.

So the interesting question concerns the relation between (1) *if A then C* and (2) *A only if C*. But consider the following two-part argument:

> Suppose we are given *if A then C*; this means that *A*'s truth implies *C*, so we will only have *A* if *C* obtains as well – i.e. *A only if C* holds.
>
> Conversely, suppose we are given *A only if C*; this means that if *A* is true then, willynilly, we get *C* as well, i.e. *if A then C*.

It seems then that (1) and (2) imply each other. So, plausibly,

> In many cases, propositions of the form *if A then C* and *A only if C* are equivalent.

To take a simple example, consider the following pair:

(3) If Einstein's theory is right, space-time is curved.
(4) Einstein's theory is right only if space-time is curved.

These two do indeed seem equivalent. (As so often, however, ordinary language has its quirks: so can you think of examples where a proposition of the form *if A then C* seems *not* to be straightforwardly equivalent to the corresponding *A only if C*?)

(b) Concentrate on cases where *A only if C* and *if A then C* do come to the same. Then if we want to render them into a formal language with only truth-functional connectives, we will have to translate them the same way, by using the material conditional. So in particular, *A only if C* will be rendered by the corresponding $(\alpha \to \gamma)$, where α is the formal translation of *A*, and γ the translation of *C*.

Do be careful: translating 'only if' conditionals gets beginners into a tangle surprisingly often. But the basic rule is simple : first put an 'only if' conditional into the form *A only if C* and then replace the 'only if' with the arrow '→', to get the corresponding $(\alpha \to \gamma)$.

(c) For a simple example, consider the following argument:

I Jack came to the party only if Jo did. Also Jill only came to the party if Jo did. Hence Jo came to the party if either Jack or Jill did.

(Note how the 'only' can get separated in ordinary English from its associated 'if'.)

This argument is evidently valid. What about its correlate in a PLC language? Using 'P' to render the claim *Jack came to the party*, 'Q' for *Jill came to the party*, and 'R' for *Jo came to the party*, we can transcribe the argument as follows:

I′ $(\mathsf{P} \to \mathsf{R}),\ (\mathsf{Q} \to \mathsf{R})\ \therefore\ ((\mathsf{P} \lor \mathsf{Q}) \to \mathsf{R})$.

A simple truth-table shows that this formal version is indeed tautologically valid:

P	Q	R	$(P \to R)$	$(Q \to R)$	$((P \lor Q) \to R)$
T	T	T			T
T	T	F	F		F
T	F	T			T
T	F	F	F		F
F	T	T			T
F	T	F	T	F	F
F	F	T			T
F	F	F			T

(d) Finally, putting things together, what about the so-called *biconditional*, i.e. *A if and only if C*?

Plainly, this two-way conditional says more than either of the separate one-way conditionals *A if C* and *A only if C*. Which doesn't stop careless authors sometimes writing *A if C* when they really mean *A if and only if C*. And it doesn't stop careless readers understanding the one-way *A if C* as the two-way *A if and only if C* even when only the one-way conditional is intended. So do take care!

The biconditional *A if and only if C* is the conjunction of *A if C* and *A only if C*. The second conjunct we have just suggested should be translated into our formal language by the relevant $(\alpha \to \gamma)$. The first conjunct, equivalently *If C, then A*, is rendered by $(\gamma \to \alpha)$. Hence the whole biconditional can be rendered by $((\gamma \to \alpha) \land (\alpha \to \gamma))$.

Now, that is a just a little cumbersome. So it is quite common to introduce another new symbol '↔' for the two-way truth-functional conditional where, by definition,

> $(\alpha \leftrightarrow \gamma)$ is tautologically equivalent to $((\gamma \to \alpha) \land (\alpha \to \gamma))$.

This is the so-called *material biconditional*; and a simple calculation shows that its truth-table will be as displayed.

α	γ	$(\alpha \leftrightarrow \gamma)$
T	T	T
T	F	F
F	T	F
F	F	T

You really should know about this biconditional truth-function and its most basic properties, and be able to recognize symbols for it when you see them elsewhere. However, to reduce clutter, we will *not* officially deploy '↔' in the main text of this book; instead, we will relegate its use to a series of exploratory examples in various of the end-of-chapter Exercises.

(e) We should add, by the way, that as well as the *formal* symbol for the biconditional, there is a widely used *informal* abbreviation that you should certainly know about:

> Logicians and logically-inclined philosophers often write the ordinary-language biconditional *A if and only if C* as simply *A iff C*.

17.6 PLC syntax and semantics, officially

The material conditional – like any truth-functional way of combining wffs – can be expressed using the existing resources of a PL language. It is easily seen that both

$(\neg\alpha \lor \gamma)$ and $\neg(\alpha \land \neg\gamma)$ have the same truth-table as $(\alpha \to \gamma)$. (Check this! We return to these equivalences in §18.2.)

So, as we said before, we certainly do not *need* to add a new basic symbol for the material conditional. However, there are trade-offs between various kinds of simplicity here. On the one hand, if we add a symbol for the conditional, we thereby get nicer-looking 'translations' (if we can call them that). On the other hand, in complicates our story a little: e.g. in developing formal apparatus for regimenting proofs involving these connectives, we will need additional inference rules to govern the new connective. On balance, however, it is worth making the trade: so we will continue to consider languages which add the new connective for the material conditional.

(a) Languages which have the truth-functional conditional built in alongside the other connectives are our PLC languages. Here, then, is a summary of the official syntax of such languages (compare §9.1):

The alphabet of a PLC language is that of a PL language with '→' added.

The definition of atomic wffs remains the same as for PL languages.

The rules for forming PLC wffs are now

(W1) Any atomic wff counts as a wff.
(W2) If α and β are wffs, so is $(\alpha \land \beta)$.
(W3) If α and β are wffs, so is $(\alpha \lor \beta)$.
(W4) If α and β are wffs, so is $(\alpha \to \beta)$.
(W5) If α is a wff, so is $\neg\alpha$.
(W6) Nothing else is a wff.

Our previous result about the uniqueness of parse trees for wffs, our previous definition of the idea of a main connective, etc., will all carry over. We need not pause over such matters.

(b) One comment, however, on the particular choice of symbolism for the material conditional.

Most logicians now use '→'. But you will also often encounter the alternative old-school symbol '⊃', due to the nineteenth century mathematician Giuseppe Peano and popularized by Bertrand Russell (and '≡' is an old alternative for the material biconditional). Indeed, there might perhaps be something to be said for keeping to the old notation when the truth-functional material (bi)conditional is explicitly intended. But we will follow modern notational practice.

(c) As for the semantics of PLC, it is a highly contentious question just how well the material conditional captures the core *meaning* of even simple ordinary 'if's. We will say more about this in the next chapter.

However, it is entirely straightforward to extend the story about *valuations* for PL languages to cover PLC languages. We can now sum up how '→' affects valuations by adding the truth-table for this new connective to our previous panel:

The truth tables for the PLC *connectives:*

α	β	$(\alpha \land \beta)$	$(\alpha \lor \beta)$	$(\alpha \to \beta)$
T	T	T	T	T
T	F	F	T	F
F	T	F	T	T
F	F	F	F	T

α	$\neg\alpha$
T	F
F	T

These tables get applied in evaluating molecular wffs exactly as before. Further, as we have already seen in §17.4, the ideas of being a tautologically valid inference, etc., can again be simply carried over to cover arguments in PLC languages.

17.7 '$\to$' versus '$\vDash$' and '$\therefore$'

(a) Recall that in §16.3 we drew a very sharp distinction between the object-language inference marker '$\therefore$' and the metalinguistic sign '$\vDash$' for the entailment relation. We now have another sign which we must also distinguish very clearly from *both* of these, namely the object-language conditional '$\to$'. Let's finish this chapter by sorting things out carefully.

As we stressed before,

(1) $(\mathsf{P} \land \mathsf{Q}) \therefore \mathsf{Q}$

is a mini-argument expressed in a PL or PLC language. Someone who asserts (1), assuming that the atoms have been given some content, is asserting the premiss, asserting the conclusion, and indicating that the second assertion is derived from the first one. Compare, for example, the English *Jack is a physicist and Jill is a logician. Hence Jill is a logician*.

By contrast

(2) $((\mathsf{P} \land \mathsf{Q}) \to \mathsf{Q})$

is a single proposition in a PLC language. Someone who asserts (2) asserts the whole material conditional in that language (whose content will depend on the content we have given the atoms); but they of course assert neither the antecedent '$(\mathsf{P} \land \mathsf{Q})$' nor the consequent '$\mathsf{Q}$'. Compare the English *If Jack is a physicist and Jill is a logician, then Jill is a logician*.

By contrast again

(3) $(\mathsf{P} \land \mathsf{Q}) \vDash \mathsf{Q}$

is another single proposition, but this time it is a proposition in our extended English metalanguage, and the wffs here are not used but mentioned. For (3) just abbreviates the English

(3′) '$(\mathsf{P} \land \mathsf{Q})$' tautologically entails 'Q'.

Compare *'If Jack is a physicist and Jill is a logician' logically entails 'Jill is a logician'*.

(b) There's an unfortunate practice that – as we said – goes back to Russell of talking, not of the 'material conditional', but of 'material *implication*', and reading something of the form '$(\alpha \to \beta)$' as α *implies* β. If we also read '$\alpha \vDash \beta$' as α *(tauto)logically implies* β, this makes it sound as if '$(\alpha \to \beta)$' is the same sort of claim as '$\alpha \vDash \beta$', only the first is a weaker version of the second. But, as we have just emphasized, these are in fact claims of a quite different kind, one in the object language, one in the metalanguage. Talk of 'implication' can blur the very important distinction. (Even worse, you will often find e.g. '$\Rightarrow$' being used in informal discussions to mean either '$\to$' or '$\vDash$', and you have to guess from context which is intended. Never follow this bad practice!)

(c) However, we can link 'logical implication' in the sense of entailment to 'material implication' in the sense of the material conditional by the following result:

> Let α, β be wffs from a PLC language: then $\alpha \vDash \beta$ if and only if $\vDash (\alpha \to \beta)$.

In plain words: a PLC argument of the form $\alpha \therefore \beta$ is tautologically valid if and only if (iff) the corresponding material conditional $(\alpha \to \beta)$ is a PLC tautology. Why so?

> By definition, for any wffs α and β, $\alpha \vDash \beta$ iff there is no assignment of values to the atoms which appear in the wffs which makes α true and β false, i.e. which makes $(\alpha \to \beta)$ false. Hence $\alpha \vDash \beta$ iff every assignment of values to the relevant atoms makes $(\alpha \to \beta)$ true, i.e. $\vDash (\alpha \to \beta)$.

(d) By similar reasoning, we can show that $\alpha_1, \alpha_2, \ldots, \alpha_n, \beta \vDash \gamma$ if and only if $\alpha_1, \alpha_2, \ldots, \alpha_n \vDash (\beta \to \gamma)$, for any αs. This is a formal analogue of the following principle governing reasoning to a conditional conclusion:

> Suppose, using some background premisses, we can argue from the additional temporary assumption B to the conclusion C. Then, keeping those background premisses fixed, we can infer from them that *if B then C*.

We will return to discuss this important rule of inference – so-called *conditional proof* – in Chapter 21.

17.8 Summary

> The only truth-functional connective that is a candidate for translating the conditional of ordinary discourse is the so-called material conditional. If we represent this connective by '$\to$', then $(\alpha \to \gamma)$ is false just when α is true and γ is false, and is true otherwise.
>
> Since PL languages are expressively complete, we don't need to add new a symbol to such languages to express the material conditional. But it is convenient to do so. The extended languages, with the symbol '$\to$' added in the obvious way, will be termed PLC languages.
>
> At least some ordinary-language arguments using conditionals can be rendered into PLC languages using '$\to$' while preserving their intuitive validity or invalidity. However, the exact relation between the general run of ordinary 'if's and '$\to$' is contentious, as we will see.

Many conditionals for the form *A only if C* are equivalent to the corresponding conditional of the form *if A then C*; so both sorts of conditional can be rendered into PLC languages using '$\to$' equally well (or equally badly!).

It is imperative to sharply distinguish '$\to$' from both '$\therefore$' and '$\vDash$'.

Exercises 17

(To be added)

18 'If's and '→'s

We saw in the last chapter that the logical behaviour of '→' in PLC languages is parallel to that of the vernacular 'if', in at least *some* cases, in *some* key respects. But just how close is the relation between ordinary conditionals and material conditionals?

18.1 Types of conditional

(a) We are discussing the biodynamics of kangaroos (as one does). You say:

(1) If kangaroos had no tails, they would topple over.

How do we decide whether you are right? By imagining a world very like our actual world, with the same physical laws, and where kangaroos are built much the same except for the lack of tails. We then work out whether the poor beasts in such a possible world would be unbalanced and fall on their noses.

In short, we have to consider not the actual situation (where of course kangaroos *do* have tails), but other possible ways things might have been. We have to engage in counterfactual reasoning. So let's call a conditional that invites this kind of evaluation (i.e. evaluation by thinking not about the world as it is but about other ways things might have been) a *counterfactual conditional* or a *possible-world conditional*.

(b) Here is a memorable pair of examples:

(2) If Oswald didn't shoot Kennedy in Dallas, someone else did.
(3) If Oswald hadn't shot Kennedy in Dallas, someone else would have.

Let's assume that we agree that, in the actual world, Kennedy was definitely shot in Dallas, and we also believe that Oswald did it, acting alone. What should we then think about (2) and (3)?

Evidently, since in the actual world Kennedy was shot, someone must have done it. Hence, if (to our surprise) not Oswald, someone else. So we'll take (2) to be true.

But to decide whether (3) is true, we have to consider a non-actual possible world, a world like this one except that Oswald missed. Keeping things as similar as we can to the actual world (as we believe it to be), Oswald would still have been acting alone. There would still be no back-up marksmen. In such a possible situation, Kennedy would have left Dallas unscathed. Hence, we'll take (3) to be false.

Since they take different truth-values, (2) and (3) must have different contents. Which reinforces the intuition that a 'would have' possible-world conditional like (3) means something different from a simple 'did' conditional like (2).

(c) Now, the logic of possible-world conditionals is beyond the scope of this book, but one thing is clear:

> Possible-world conditionals are not truth-functional, and can not be adequately translated by the material conditional.

For a material conditional has its actual truth-value fixed by the actual truth-values of its constituent sentences. By contrast, the truth-value of a possible-world conditional depends on what happens in other possible scenarios; so its truth-value can't be fixed just by the this-worldly values of the antecedent and consequent. Hence, entirely uncontroversially, the truth-functional rendition can at most be used for conditionals like (2) and not for those like (1) or (3).

(d) What is controversial, though, is the range of conditionals that are appropriately 'like (2)' and which are not 'like (1) or (3)'.

Conditionals grammatically like (1) and (3) are conventionally called *subjunctive* conditionals, for supposedly they are couched in the subjunctive mood (though this is often contested). By contrast, (2) – and all the conditionals in examples **A** to **I** at the beginning of the last chapter – are conventionally called *indicative* conditionals, being framed in the indicative mood. So does the traditional grammatical subjunctive/indicative distinction mark the distinction between the possible-world conditionals which are definitely not truth-functional and the rest?

Arguably not, as we will see in the Appendix. But we don't want to get bogged down over contentious issues about the classification of conditionals any further here. From now on, then, let's leave aside possible-world conditionals (whichever exactly those are), and concentrate on core cases of 'indicative' conditionals.

We will also set aside conditionals like 'If you heat an iron bar, it expands' or 'If a number is even and greater than two, then it is the sum of two primes' (although these are in the indicative mood). For such conditionals are in fact generalizations, equivalent to e.g. 'Take any iron bar, if you heat it, it expands' or 'Any number is such that, if it is even and greater than two, then it is the sum of two primes'. So we narrow our focus further, and consider just *singular*, i.e. ungeneralized, indicative conditionals. Let's ask: how well does the material conditional work in capturing the logical content of at least *these* conditionals ('simple' ones, as we arm-wavingly called them)?

18.2 Simple conditionals as truth-functional: for

We saw how the singular indicative conditionals in arguments **A** to **I** from the last chapter can be rendered into PLC using the truth-functional material conditional in a way that preserves facts about validity and invalidity. Here then is a hopeful proposal:

> (*if* = →) '→' does stand to some 'if's as '∧' stands to 'and' and '∨' stands to 'or'; the material conditional captures the core logical role of those 'if's which feature in singular indicative conditionals.

In this section, we briskly argue *for* this attractively easy position; then in the next section we give an even brisker argument *against* it.

(a) How does an ordinary-language simple (i.e. singular indicative) conditional *if A then C* relate to the propositions *either not-A or C* and *it isn't the case that both A and not-C*? Here is a pair of two-part arguments (think of them as 'If/And' and 'If/Or'):

IA (i) The claim *if A then C* rules out having *A* true and *C* false. So *if A then C* implies *it isn't the case that both A and not-C*. (This is just the falsehood condition (FC) which we met in §17.2.)

(ii) Conversely, suppose we are given that *it isn't the case that both A and not-C*. Then we can infer that if *A* is actually true we can't have *not-C* as well: in other words *if A then C*.

IO (i) Suppose *if A then C*. So we either have *not-A*, or we have *A* and hence *C*. So *if A then C* implies *either not-A or C*.

(ii) Conversely, suppose we are given *either not-A or C*. Then if not the first, then the second. So we can infer *if A then C*.

The two-part argument **IA** implies that *if A then C* is equivalent to the truth-functional *it isn't the case that both A and not-C*. The two-part argument **IO** implies that *if A then C* is equivalent to the truth-functional *not-A or C*. Therefore – using α, γ to stand in for formal translations of the English clauses A, C – we can render the truth-relevant content of indicative conditionals *if A then C* by something of the form $\neg(\alpha \land \neg\gamma)$ and equally by something of the form $(\neg\alpha \lor \gamma)$.

But, by trivial truth tables, both those are of course tautologically equivalent to $(\alpha \to \gamma)$, as we noted in passing at the beginning of §17.6. So we can indeed render the core content of simple conditionals using '$\to$'.

(b) That argument for (*if* $= \to$) is indeed brisk but it is entirely abstract. So let's amplify one of the steps by giving a realistic concrete illustration of what looks like an equivalence between something of the form *if A then C* and the corresponding *not-A or C* – or, what comes to the same, between *if not-B then C* and the corresponding *B or C*.

Suppose I vividly remember that Jack and Jill threw a party together last Easter, but can't recall whose birthday it was celebrating. On this basis, I believe

(1) Either Jack has a birthday in April or Jill does.

Now, you tell me – a bit hesitantly – that you think that maybe Jack has an October birthday. I agree that you might, for all I know, be right; but I will still infer from (1) that

(2) If Jack was not born in April, then Jill was.

In the context, deriving (2) is surely perfectly acceptable. And note, it doesn't commit me for a moment to supposing that there is any causal or other intrinsic connection between the facts about Jack and Jill's birth months. I just think that at least one of the propositions *Jack was born in April* and *Jill was born in April* happens to be true, and thus if not the first, then the second.

So (1) implies (2); and since (2) unproblematically implies (1), these two therefore are equivalent in the sense of being interdeducible. Hence perfectly acceptable conditionals like (2) can be no-more-than-material in what they commit us to about the world.

(c) Of course, when we assert an ordinary conditional, we often *do* think that there is e.g. a causal connection between the matters mentioned in the antecedent and the consequent ('If ice was applied, the swelling went down'). But equally, when we assert a disjunction, it is quite often because we think there is some mechanism ensuring that one disjunct or the other is true ('Either the e-mail was sent straight away or you will have received a warning message that the e-mail is queued'). However, even if our *ground* for asserting that disjunction is our belief in an appropriate mechanism, what we actually *say* is true just so long as one or other disjunct holds. Likewise, our *ground* for asserting a conditional may be e.g. a belief in some mechanism that ensures that if the antecedent holds the consequent does too. But the 'birthdays' example shows that such a causal mechanism doesn't have to be in place for a conditional to be true.

18.3 Another case for equating 'if' and '→'

We should next note a very important family of cases where simple indicative 'if's apparently do behave just like '→'s (at least as far as truth conditions are concerned).

Consider again Goldbach's Conjecture, the famous – and still unproved – proposition that

(1) If a number is even and greater than two, then it is the sum of two primes,

(where 'number' means 'positive integer'). Now, this is of course not a straightforward conditional of the form *if A then C*, with *A* and *C* whole propositions. Rather, it is a *generalized* conditional, equivalent to

(2) Any number is such that, if it is even and greater than two, then it is the sum of two primes

The truth of (1) therefore requires the truth of *every* particular instance of

(3) If *n* is even and greater than two, then *n* is the sum of two primes,

where '*n*' gets replaced in turn by '1', '2', '3', ... Now ask the key question *what kind of 'if' is involved in these particular instances, given that every single instance must be true*?

Take first the instances where the antecedent of the conditional is *false*. These are instances where *n* is either 2 or is an odd number – and in these cases, the consequent of the conditional may take either truth-value. But either way, the whole conditional has to be true. Goldbach's Conjecture can't be falsified by what happens to odd numbers! Hence, when the antecedent of an instance of (3) is false, for this instance to come out true, we need the conditional to behave just like a material conditional: we need the conditional to come out true irrespective of whether the consequent is true or false.

Take now the remaining instances of (3), where the antecedent of the conditional is *true*. So we are dealing with a number *n* which *is* an even number great than two. Then, we surely only need the consequent to be true, i.e. only need *n* to be the sum of two primes, for the whole conditional to be true. And should the consequent be false, and *n* not be the sum of two primes, then the whole conditional will be false. In other words, when the antecedent of an instance of (3) is true, the conditional in this instance again behaves like a material conditional.

In short: it is natural to suppose that *all* the particular instances of (at least some) generalized conditionals, like Goldbach's Conjecture (1), do behave *just* like material conditionals.

18.4 Simple conditionals as truth-functional: against

So far, then, so favourable for the proposal (*if* = →). Now for an argument that goes flatly in the opposite direction.

In fact, there are a number of cases where treating ordinary indicative conditionals as material conditionals leads to strongly counter-intuitive claims about validity. But we will concentrate here on a central type of case (for more examples, see the Exercises). Compare, then, the following three arguments – where, for vividness, we temporarily borrow '→' to express the material conditional truth-function in English:

(1) Bacon didn't write *Hamlet*. So, either Bacon didn't write *Hamlet* or he was a bad dramatist.

(2) Bacon didn't write *Hamlet*. So, (Bacon wrote *Hamlet* → Bacon was a bad dramatist).

(3) Bacon didn't write *Hamlet*. So, if Bacon wrote *Hamlet*, then he was a bad dramatist.

(1) is trivially valid (as is any inference of the form *not-A, so either not-A or C* for inclusive 'or'). By definition, the conclusion of (2) is just another way of writing the conclusion of (1), and hence (2) must be trivially valid too. By contrast, the inference in (3) looks absurd. (3)'s conclusion, it will be said, is quite unacceptable (being the author of *Hamlet* makes you a good dramatist, if anything does). So how can the apparently absurd conclusion validly follow from the sensible premiss?

Generalizing, the inference pattern

(M) *not-A*, so *not-A or C*

is unproblematically and trivially reliable. So too, borrowing the arrow again, is the inference

(M′) *not-A*, so $A \to C$.

On the other hand, many inferences of the type

(V) *not-A*, so *if A then C*.

strike us as quite unacceptable. For example, we surely can *not* correctly argue e.g. 'I won't buy a lottery ticket; hence if I buy a ticket I will win', or argue 'Jill has not got a broken leg. So if Jill has a broken leg, she will go skiing today', or argue 'It's not freezing cold. So if it's freezing cold, it's boiling hot.' And so on.

Yet (*if* = →) equates the content of the vernacular conditional *if A then C* in (V) with the material conditional expressed in English by *not-A or C* in (M) – which implies that the two inference patterns should after all be exactly on a par. So inferences of the type (V) are as reliable as inferences of type (M). Which looks absurd.

18.5 Three responses

On the one hand, §§18.2 and 18.3 offer strong-looking arguments in favour of the view that simple conditionals are equivalent to material conditionals. On the other hand, §18.4 exposes what seems to be an entirely unwelcome upshot of this view. We have a problem! What to do?

We consider two initial responses which try to rescue the proposal (*if* = →) as it stands; and then we offer a third response which suggests a friendly amendment to that proposal.

(a) The first response simply asks us to think again and then to revise our initial antipathy to (V):

> Suppose I'm entirely confident that Bacon didn't write *Hamlet*. Then, I'll happily assert 'If Bacon wrote *Hamlet*, then I'm a Dutchman', 'If Bacon wrote *Hamlet*, then pigs can fly', 'If Bacon wrote *Hamlet*, then Donald Trump wrote *Pride and Prejudice*'. In the same spirit, we can also conclude 'If Bacon wrote *Hamlet*, then . . . ' for any other completion, however silly the result. Why not?

But the trouble with this retort is that 'Dutchman' conditionals – as we can call them – strike us as jokey idiom. To say 'If Bacon wrote *Hamlet*, then I'm a Dutchman' seems too much like going through a *pretence*, playing at asserting a conditional as a dramatic way of inviting you to draw the obvious modus tollens conclusion that Bacon didn't write *Hamlet*. Can we safely erect a theory of conditionals on the basis of exceptional outliers like this?

(b) The second response we will consider is more conciliatory, allows that (V) not only looks absurd at first sight but continues to do so on further reflection, but still tries to reconcile this with (*if* = →):

> It would, we can agree, be uncooperative of me to assert something equivalent to $A \to C$ when I already know that A is simply false. For I could much more informatively simply tell you, straight out, that *not-A*. That's why you can normally expect me to have grounds for asserting $A \to C$ other than belief in the falsity of the antecedent.
>
> Similarly, assuming (*if* = →), for the equivalent ordinary-language proposition *if A then C*. Even though it is really a material conditional, you'll reasonably expect me to have some grounds for asserting it other than belief in *not-A*. Because of that strong expectation, inferences of the form (V) will strike us as peculiar, even though strictly speaking correct.

But does this response do enough? Yes, it would usually be oddly unhelpful of me to assert a bare disjunction when I believe e.g. that the first disjunct is definitely true. So, yes, in many contexts, it would be odd simply to assert the disjunction *not-A or C*, or equivalently assert $A \to C$, when my reason is I think that *not-A*. However, granted all that, nothing strikes us as odd about the explicit inference (M) *not-A; so not-A or C*. Likewise, there is nothing odd about the equivalent inference (M′) *not-A; so* $A \to C$, once we understand the symbol. That's because in this case the default presumption that

I might have grounds for $A \to C$ (i.e. *not-A or C*) other than belief in *not-A* is plainly being *cancelled*: I am here quite frankly and explicitly offering *not-A* as my grounds!

In sum, there its nothing odd about (M′), even allowing for what can be expected from co-operative conversationalists. And if the ordinary conditional is merely the material conditional again, then there should equally be nothing odd about (V), still allowing for what can be expected from co-operative conversationalists. Yet surely the cases *are* different.

(c) It still seems, then, that an unqualified identification of 'if' and the material conditional can't explain the radical difference in plausibility between inferences (M) and (V). So that strongly suggests that, even concentrating on simple cases, there is *something* more to the meaning of 'if' than a merely material conditional. But what? Here is a new outline proposal:

> Recall our brief discussion in §8.2(b) about the difference between *A but B* and *A and B*. We suggested that these claims don't differ in what it takes for them to be true. Rather, the contrast between 'but' and the colourless 'and' is (very roughly) that *A but B* is typically reserved for use when the speaker presumes that there is some kind of contrast between the truth or the current relevance of *A* and of *B*.
>
> Well, could it be similar for claims for the form *if A then C* and $A \to C$? In other words, perhaps these don't differ in what it takes for them to be true. But still, there *is* a difference between 'if' and the colourless material conditional. In this case, *if A then C* is reserved for use when the speaker is – as it were – promising to endorse a modus ponens inference to the conclusion *C* should it turn out that *A* is true.

This seems a friendly amendment of the bald equation (*if* = $\to$). For it sticks to the core idea that the assertions *if A then C* and $A \to C$ (i.e. *not-A or C*) come to same as far their logic-relevant truth-conditions are concerned. That explains why the arguments **IA** and **IO** in §18.2 look compelling. But we also get an explanation for why the inference from *not-A* to *if A then C* typically seems quite unacceptable (even though truth-preserving).

How so? The plausible suggestion is that the use of 'if' signals that the conditional it expresses can be used in an inference to its consequent. Now, suppose that I start off by believing *not-A*. Then, I will accept *not-A or C* or equivalently $A \to C$. But of course, I *won't* in this case be prepared to go on to infer *C* should I later come round to accepting *A*. Instead, because I have changed my mind about the truth-value of *A*, I will now *take back* my earlier endorsement of $A \to C$. In other words, if my only reason for accepting $A \to C$ is a belief in *not-A*, I won't be prepared to go on to endorse using the conditional in a modus ponens inference to the conclusion *C*. And that is why, according to our suggestion about the role of 'if', I won't be prepared to assert *if A then C* just on the ground that *not-A*.

Which is a neat story. But is it right?

18.6 Adopting the material conditional

We are going to have to leave that last question hanging. The debates in the vast literature on conditionals are sophisticated and intriguing but frustratingly inconclusive. We have only scratched the surface, and we certainly can't review the ramifying arguments any further here (though for some additional introductory remarks, see the Appendix 'More on conditionals').

However, for our purposes in this book, we can perhaps happily side-step all these debates. Let's finish this chapter by explaining how.

(a) Suppose we decide that the behaviour of everyday 'if' is too complex to be accommodated by a simple and obvious theory (that's what the inconclusive philosophical literature about conditionals might seem to show). Well, in this case, should we even be *trying* to reflect those complexities in a formal language?

As we have seen, rendering arguments involving conditionals into a formal language by using the material conditional can often get things right in straightforward cases – i.e., get things right as far as the resulting verdicts about the validity or invalidity of the argumen oncerned. So changing tack, perhaps we should be content with that. In other words, perhaps we should instead regard the material conditional as a suitable partial *substitute* for the vernacular conditional, one which is sufficiently conditional-like to be happily adopted for some important formal purposes, while being elegantly clear, perfectly understood, and very easy to work with.

This is in fact the line typically taken by mathematicians who want a useful and easily-managed logic for regimenting their arguments. Indeed, this is exactly how the truth-functional conditional was introduced by Frege, the founding father of modern logic, in his *Begriffsschrift*. Frege's aim was to provide a setting in which mathematical reasoning, in particular, could be reconstructed entirely clearly and unambiguously – and for him, such clarity requires departing significantly from what he calls "the peculiarities of ordinary language". So he introduces a formalized framework which then takes on a life of its own. Choice of notation apart, the central parts of Frege's formal apparatus including his truth-functional conditional, together with his basic logical principles (bar one), turn out to be exactly what mathematicians need.

That's why modern mathematicians – who do widely use logical notation for clarificatory purposes – often introduce the material conditional '→' in text books as part of their 'mathematical English', and then cheerfully say (in a Fregean spirit) that this tidy notion is what *they* are going to mean by 'if'. It serves them perfectly in regimenting their theories, e.g. in giving their formal theories of arithmetic or set theory. As we saw, it works well, in particular, in regimenting general claims like 'if a number is even and greater than two, then it is the sum of two primes'. And, as we will see in Chapter 21, the inference rules that the material conditional obeys are just the rules that mathematicians and others already use by default in reasoning with conditionals. So '→' is sufficiently conditional-like for their purposes, even if – contrary to the hopeful story in the last section – it doesn't capture all the truth-relevant content of ordinary 'if's.

This gives us, then, more than enough reason to continue exploring what happens when we adopt '→' as a 'clean' substitute for the conditional in our formal languages – a substitute which serves many of the central purposes for which we want conditionals, especially in mathematical contexts.

(b) However, a word of warning. If the material conditional perhaps does not always capture all the content of vernacular conditionals, if we are thinking of it more as a substitute apt for use in some (but not all) contexts, then we'll have to be correspondingly cautious in using the sort of two-stage procedure for evaluating ordinary-language arguments that we illustrated in §17.4.

The idea, recall, is that (1) we render a vernacular argument involving conditionals into an appropriate PLC language, and then (2) we evaluate the corresponding PLC argument by running the truth-table test. But we now have to ask ourselves, case-by-case, whether the rendition of the vernacular argument into PLC using the truth-functional conditional at stage (1) really hits off enough of the conditional content of the premisses and conclusion for the verdict on the PLC version to carry back to a verdict on the original argument. So our motto has to be: *proceed with caution*.

18.7 Summary

As we saw in the previous chapter, a good number of ordinary-language arguments using conditionals can be rendered using '→' while preserving their intuitive validity/invalidity. However this at best works only for some singular 'indicative' conditionals, and not for so-called 'subjunctive/counterfactual' (or 'possible world') conditionals.

There are serious issues about how far, even in the best cases, '→' can capture all of the core meaning of everyday 'if'. Identifying the two leads to oddities like the acceptability of arguments of the form *Not-A; so if A, then C*.

However, the material conditional can still be recommended as a perfectly clear *substitute* for the ordinary language conditional, with a perfectly determinate logic, suitable for use by mathematicians and others.

Also see the Appendix 'More on conditionals'.

Exercises 18

(To be added)