2 Possible Worlds

2.1 The possible-worlds analysis of possibility and necessity

An important breakthrough in the history of modal logic was the development of "possible-worlds semantics" in the 1940s-60s. The central idea of possible-worlds semantics is to analyze modal notions in terms of truth at possible worlds. In its simplest form, the analysis goes like this:

> A proposition is possible iff it is true at some possible world.
> A proposition is necessary iff it is true at all possible worlds.

In philosophy jargon, a **possible world** is a maximally specific possibility. An example of a possible world is the **actual world** – the totality of everything that is the case. In the actual world, light travels faster than sound and the Conservatives won the 2019 UK election. In other possible worlds, Labour won the election. In yet others, sound travels faster than light.

The possible-worlds analysis translates modal statements into quantificational statements about possible worlds. You may feel uneasy about this. Talking about worlds other than the actual world may strike you as fanciful and unscientific. Besides, you may wonder if anything is really gained by the translation, since we now face the question of what sorts of worlds should be classified as "possible".

Remember that there are different flavours of modality. A proposition might be epistemically possible, historically possible, metaphysically possible, and so on. If we want to analyse all these kinds of possibility in terms of possible worlds, we need different flavours of worlds. There must be epistemically possible worlds, historically possible worlds, metaphysically possible worlds, etc. And if we ask how these types of worlds are defined it looks like we have to turn back to relevant features of the actual world. The ultimate reason why you can't go from Auckland to Sydney by

train is surely that there is no suitable train line here in our world, not that you don't make the journey in some non-actual worlds.

These objections cast doubt on the possible-worlds analysis as a piece of reductive metaphysics. But the metaphysics of modality is not our topic. When we use the possible-world analysis, we don't assume that the translation in terms of possible worlds reveals the metaphysical grounds of the original modal statements. We merely assume that the original statements can be paraphrased in the fanciful language of possible worlds.

For a first glimpse of why this might be useful, consider the following hypothesis.

$$\Box\Diamond\Box p \models \Box p$$

Is this true? If something is necessarily possibly necessary, does it follow that it is necessary? Hard to say. We know that A entails B iff there is no conceivable scenario in which A is true and B false, under any interpretation of the non-logical expressions. The problem is that it is not obvious what a scenario would have to look like for $\Box\Diamond\Box p$ to be true, under a given interpretation of p.

The possible-worlds analysis can help clear things up. By the possible-worlds analysis, $\Box\Diamond\Box p$ says that $\Diamond\Box p$ is true at every possible world. Since we want to know in which scenarios $\Box\Diamond\Box p$ is true, we will assume that every scenario contains a whole range of possible worlds. The hypothesis that $\Box\Diamond\Box p$ is true in a scenario now reduces to the hypothesis that $\Diamond\Box p$ is true at every world in the scenario. $\Diamond\Box p$ says that $\Box p$ is true at some world. So if $\Diamond\Box p$ is true at every world in a scenario then $\Box p$ is true at some world in the scenario. And if $\Box p$ is true at some world in a scenario then p is true at every world in the scenario. This is just what $\Box p$ says. So whenever $\Box\Diamond\Box p$ is true in a scenario (under some interpretation of p), then $\Box p$ is true in that scenario (under that interpretation). $\Box\Diamond\Box p$ entails $\Diamond p$.

Exercise 2.1

Explain, in the same informal manner, why $\Diamond p$ does not entail $\Box p$, assuming the possible-worlds analysis of the box and the diamond.

2.2 Models

In section 1.4, I defined validity and entailment in terms of scenarios and interpretations. A sentence is valid, I said, iff it is true in every conceivable scenario under every interpretation of the non-logical expressions. This is a little vague. What, exactly, is a conceivable scenario, and what counts as a relevant interpretation? Also, scenarios and interpretations are unwieldy objects. It is difficult to give a full description of a scenario and an interpretation. If all we care about is which $\mathfrak{L}_M$-sentences are true and which are false in a scenario under a particular interpretation, most of the details turn out to be irrelevant. This observation will lead the way to a more precise definition of validity and entailment.

Suppose I tell you the following about a scenario S and an interpretation I of the sentence letters.

> There are three worlds in S, w_1, w_2, and w_3. Under the interpretation I, the sentence p expresses a proposition that is true at w_1, false at w_2, and true at w_3. All other sentence letters express propositions that are false at all three worlds.

This tells you almost nothing about what the scenario looks like. You don't know if w_1 is a world at which it is currently raining, or who won which elections at w_2. You also don't know what the sentences letter mean under my interpretation. Does p mean that it is raining? That Labour won the 2019 election? I haven't told you. Yet the sparse information I have given is enough to determine the truth-value of every $\mathfrak{L}_M$ sentence at every world.

Exercise 2.2

Which of the following sentences are true at w_1 in my scenario S under my interpretation I?

(a) $\neg p$
(b) $\neg p \rightarrow \Box p$
(c) $\Box p$
(d) $\Diamond \Box p$
(e) $\Diamond \Diamond p \vee \Diamond \Box p$
(f) $\Box(\Box p \rightarrow p)$

A joint representation of a scenario and an interpretation (of non-logical expressions) that contains just enough information to determine the truth-value of every sentence is called a **model**. Just as a model airplane often leaves out important aspects of a real airplane – the motor, the seats, etc. – models in logic leave out many important aspects of the scenarios and interpretations they represent.

Adopting the simple possible-worlds analysis of the box and the diamond, we can define a model for $\mathfrak{L}_M$ as consisting of two parts. First, there is a set of things we call "worlds". They don't need to be genuine worlds. They can be arbitrary (usually not further specified) objects whose job is to represent genuine worlds. Second, there is an "interpretation function" that tells us for each sentence letter at which of the "worlds" it is true.

Definition 2.1

A **basic model** of $\mathfrak{L}_M$ is a pair $\langle W, V \rangle$ of

- a non-empty set W, and
- a function V that assigns to each sentence letter of $\mathfrak{L}_M$ a subset of W.

In the next chapter, we will replace this definition by a slightly more complicated definition. That's why I've called models of the present kind 'basic'.

(If you are not familiar with elementary concepts of set theory: A *set* is a collection of objects, called the *members* or *elements* of the set. Sets can be defined by listing their members enclosed in curly braces: '$\{a, b, c\}$'. The *empty set*, with no members, is denoted by '$\emptyset$'. A *subset* of a set X is a set all whose members are members of X. A *function* is a mapping – a kind of abstract machine that takes objects of a certain kind as input and outputs objects of a possibly different kind.)

The interpretation function V in a model maps each sentence letter to the set of worlds at which the sentence is true. For example, if W contains three worlds w_1, w_2, and w_3, and $V(p) = \{w_1, w_3\}$ – meaning that V maps p to the set $\{w_1, w_3\}$ –, then p is true at w_1 and w_3 but not at w_2.

Notice that an interpretation function only specifies at which worlds the *sentence letters* are true. V is defined for p, q, and r, but not for $p \to q$ or $\Box p$ or $\Diamond\Box q$. This is the key idea behind the possible-worlds analysis. Once we know at which worlds each sentence letter is true, we have all we need to determine the truth-value of every sentence at every world.

To formally define how the truth-value of complex sentences is determined, I will use (meta-linguistic) statements of the form

$$M, w \models A$$

as shorthand for

A is true at world w in model M.

I use '$M, w \not\models A$' for the negation of '$M, w \models A$'.

Yes, it's the same turnstile that we use for entailment and validity. This should cause no confusion because it is usually clear if the things to the left of the turnstile are $\mathfrak{L}_M$-sentences or meta-linguistic expressions for a model and a world. (In its present use, the turnstile is often pronounced 'makes true' or 'satisfies'.)

The relation $\models$ between a model, a world and an $\mathfrak{L}_M$-sentence is defined as follows.

Definition 2.2: Basic Possible-Worlds Semantics

If $M = \langle W, V \rangle$ is a basic model, w is a member of W, A is any sentence letter, and B, C are any $\mathfrak{L}_M$-sentences, then

(a)	$M, w \models A$	iff w is in $V(A)$.
(b)	$M, w \models \neg B$	iff $M, w \not\models B$.
(c)	$M, w \models B \wedge C$	iff $M, w \models B$ and $M, w \models C$.
(d)	$M, w \models B \vee C$	iff $M, w \models B$ or $M, w \models C$.
(e)	$M, w \models B \rightarrow C$	iff $M, w \not\models B$ or $M, w \models C$.
(f)	$M, w \models B \leftrightarrow C$	iff $M, w \models B \rightarrow C$ and $M, w \models C \rightarrow B$.
(g)	$M, w \models \Box B$	iff $M, v \models B$ for all v in W.
(h)	$M, w \models \Diamond B$	iff $M, v \models B$ for some v in W.

Let's go through the clauses in this definition.

Clause (a) says that a sentence letter is true at a world in a model iff the world is an element of the set of worlds which the model's interpretation function assigns to the sentence letter. This is just what I explained above.

Clause (b) says that the negation $\neg B$ of an $\mathfrak{L}_M$-sentence B is true at a world in a

model iff B is not true at that world in that model. In other words, the truth-table for negation applies locally at every world: at any world, $\neg B$ is true iff B is not true. Clauses (c)–(f) similarly tell us that the truth-tables for the other truth-functional connectives apply locally at each world.

Clauses (g) and (h) spell out the possible-worlds analysis of the box and the diamond. According to (g), a sentence $\Box B$ is true at a world in a model iff B is true at all worlds in the model. According to (h), $\Diamond B$ is true at a world in a model iff B is true at some world in the same model.

The whole definition is called a *semantics* because a semantics for a language is an account of what the expressions in the language mean, and definition 2.2 can be seen as giving the meaning of the logical expressions in $\mathfrak{L}_M$. (The non-logical expressions in $\mathfrak{L}_M$ don't have a fixed meaning.)

Since every $\mathfrak{L}_M$-sentence is built up from sentence letters with the operators covered in definition 2.2, the definition settles the truth-value of every sentence at every world in every model.

Consider, for example, the following model M:

$$
\begin{aligned}
&W = \{w_1, w_2\}\\
&V(p) = \{w_1, w_2\}\\
&V(q) = \{w_1\}\\
&V(A) = \emptyset \text{ for all other sentence letters } A
\end{aligned}
$$

This model contains only two worlds, w_1 and w_2. The interpretation function V indicates that p is true at both worlds, q is true at w_1, and all other sentence letters are true nowhere. With the help of definition 2.2, we can figure out at which of the two worlds, say, $\Box\Diamond(\Box q \to \Diamond\Box p)$ is true. We start with the smallest parts of the sentence.

1. p is true at w_1 and w_2 (by clause (a) of definition 2.2).
2. q is true at w_1 and not true at w_2 (by clause (a) of definition 2.2).
3. $\Box p$ is true at w_1 and w_2 (by 1 and clause (g) of definition 2.2).
4. $\Box q$ is true at no world (by 2 and clause (g) of definition 2.2).
5. $\Diamond\Box p$ is true at w_1 and w_2 (by 3 and clause (h) of definition 2.2).
6. $(\Box q \to \Diamond\Box p)$ is true at w_1 and w_2 (by 4, 5, and clause (e) of definition 2.2).
7. $\Diamond(\Box q \to \Diamond\Box p)$ is true at w_1 and w_2 (by 6 and clause (h) of definition 2.2).

8. $\Box\Diamond(\Box q \rightarrow \Diamond\Box q)$ is true at w_1 and w_2 (by 7 and clause (g) of definition 2.2).

Exercise 2.3

At which worlds in the model just described is $\Diamond p \rightarrow (q \vee \Diamond\Box p)$ true?

2.3 Basic entailment and validity

Using the concept of a model, we can now make the hand-wavy definition of entailment and validity from section 1.4 more precise.

Imagine a list of all conceivable scenarios and all possible interpretation of the sentence letters. By definition 1.3, a sentence is valid iff it is true in all of these scenarios under each of these interpretations. Every combination of a scenario S and an interpretation I is represented by a model that contains enough information to figure out whether any given sentence is true or false in S under I. Assuming that (conversely) every model represents some combination of a scenario and an interpretation, it follows that a sentence is valid iff it is true in every model. Similarly, some sentences Γ entail a sentence A iff A is true in every model in which all members of Γ are true.

That's the idea. There is, however, a small problem. Take a model with two worlds, $W = \{w_1, w_2\}$, and assume that $V(p) = \{w_1\}$. Is p true in this model? We can't say. Definition 2.2 only specifies under what conditions a sentence is true *at a world in a model*. We have not defined what it means for a sentence to be true in a model. So we can hardly say that a sentence is valid iff it is true in all models.

There are two ways to fix this. The conceptually cleaner response is to change the definition of a model. Intuitively, the worlds in a scenario are not all on a par. Think of a scenario in which it is raining although it might have been snowing. This scenario has worlds at which it is raining and others at which it is snowing. One of these worlds – a rain world – is special: it represents the actual world in the scenario. 'It is raining' is true in the scenario because it is raining in the actual world of the scenario. Following this line of thought, we could define a model to consist of *three* elements: a set of worlds W, an interpretation function V, and a "designated element of W" that indicates which world in W represents the actual world of the scenario. We could then say that a sentence is *true in a model* iff it is true at the actual world of

the model. Models of this type – with a designated element of W – are called *pointed models*.

We will adopt the more popular second response. Here we change the definition of entailment and validity. Instead of saying that a sentence is valid iff it is true in every model, we say that a sentence is valid iff it is true *at every world in every model*. Similarly, we say that some sentences Γ entail a sentence A iff A is true at every world in every model at which all members of Γ are true.

The two responses amount to the same thing. Since every world in every basic (un-pointed) model could be chosen as the designated world in a corresponding pointed model, a sentence is true at all worlds in all basic models just in case it is true in all pointed models. The response we adopt has the minor advantage of keeping models slightly simpler, and logicians want their models to be as simple as possible.

Definition 2.3

A sentence A is **valid** (for short: $\models A$) iff it is true at every world in every basic model.

Definition 2.4

Some sentences Γ **(logically) entail** a sentence A (for short: $\Gamma \models A$) iff there is no world in any model at which all members of Γ are true while A is false.

Exercise 2.4

Call a sentence true *throughout* a model iff it is true at every world in the model. What do you think of the following definition? $\Gamma \models A$ iff there is no model throughout which all members of Γ are true and throughout which A is false. Is this definition equivalent to definition 2.4? (Hint: consider the hypothesis that $p \models \Box p$.)

Above I mentioned an assumption implicit in our new definitions: that every model represents a pair of a conceivable scenario and interpretation. This isn't obvious. For example, if our topic is metaphysical possibility and necessity, then it may be hard to conceive of a scenario with exactly two possible worlds. Is it really conceivable that

there are only two ways the world might have been that are compatible with the nature of things? We could stipulate that a model, at least for this application, must contain at least (say) a million worlds, or infinitely many. It turns out, however, that this would make no difference to the logic. The very same sentences are valid whether we impose the restriction or not. So we'll allow for models with very few worlds. Such models are often useful as toy models to illustrate facts about entailment and validity.

2.4 Explorations in S5

By definition 2.3, a sentence is valid iff it is true at all worlds in all (basic) models. Definition 2.1 explains what a (basic) model is; definition 2.2 specifies the truth-value of any sentence at any world in any model. Together, these definitions settle which sentences are valid.

Let's start with $\Box p \to p$. This turns out to be valid. To see why, let w be an arbitrary world in an arbitrary model M. Either p is true at w or not. If p is true at w, then by clause (e) of definition 2.2, $\Box p \to p$ is also true at w. If p is not true at w, then by clause (g) of definition 2.2, $\Box p$ is not true at w in M either, and then $\Box p \to p$ is true at w by clause (e). So either way, $\Box p \to p$ is true at w. Since w and M were chosen arbitrarily, this shows that every instance of $\Box p \to p$ is true at every world in every model.

(In the last section of the previous chapter, I mentioned that for some applications of modal logic, we don't want $\Box p \to p$ to be valid. In the next chapter, we will see how this can be achieved by a slight tweak to the definitions of the present chapter.)

How about, say, $\Box p \to \Box\Box p$? If something is necessary, is it necessarily necessary? Our semantics says yes. Let w be an arbitrary world in an arbitrary model. If $\Box p$ is false at w, then $\Box p \to \Box\Box p$ is true at w, by clause (e) of definition 2.2. Suppose then that $\Box p$ is true at w. In that case, p is true at all worlds, by clause (g) of definition 2.2. And then $\Box p$ is true at all worlds, again by clause (g). And so $\Box\Box p$ is also true at all worlds, by clause (g). So whenever $\Box p$ is true at a world in a model, then so is $\Box\Box p$. By clause (e) of definition 2.2, it follows that $\Box p \to \Box\Box p$ is true at every world in every model.

Exercise 2.5

Show that $\Box p \to \Diamond p$ is valid.

There is a shorter way to show that $\Box p \rightarrow \Box\Box p$ is valid. Definition 2.2 entails that if a sentence starts with a modal operator, then its truth-value never varies from world to world. For example, if $\Diamond p$ is true at some world w in some model, then $\Diamond p$ is true at all worlds in the model. It follows that if a sentence starts with a modal operator, then its truth-value doesn't change if you stack further modal operators in front. If $\Diamond p$ is true at all worlds in a model, then so are $\Box\Diamond p$ and $\Diamond\Diamond p$.

This means that any sentence that begins with a sequence of modal operators is equivalent to the same sentence with all but the last operator removed. $\Diamond\Box\Box\Diamond\Diamond p$ is equivalent to $\Diamond p$. $\Box\Box p$ is equivalent to $\Box p$. Since replacing logically equivalent sentences inside a larger sentence never affects the larger sentence's truth-value at any world, $\Box\Box p \rightarrow \Box p$ is equivalent to $\Box p \rightarrow \Box p$. And this is obviously valid.

Do not conflate the concepts of necessity and validity. Necessity means truth at all worlds (or so we currently assume). Validity means truth at all worlds *in all models*. Whether an $\mathfrak{L}_M$ sentence is necessary generally varies from model to model. In a model whose interpretation function makes p true at all worlds, p is necessary insofar as $\Box p$ is true at all worlds. In a model whose interpretation function makes p false at some world, $\Box p$ is false at all worlds. Validity, by contrast, is not relative to a model. The sentence p is definitely not valid. The sentence $\Box p \rightarrow p$ is.

Exercise 2.6

Show that if a sentence A is valid, then so is $\Box A$.

Here is an example of an invalid sentence:

$$\Box(p \vee q) \rightarrow (\Box p \vee \Box q)$$

How could we show that this is invalid? By definition 2.3, a sentence is valid iff it is true at all worlds in all models. So we have to find some model in which there is some world at which the sentence is false. Such a model is called a **countermodel** for the sentence. The following model is a countermodel for the sentence above, as

you should verify with the help of definition 2.2.

$$W = \{w, v\}$$
$$V(p) = \{w\}$$
$$V(q) = \{v\}$$

I haven't explained at which worlds sentence letters other than p and q are true, because it doesn't matter.

Exercise 2.7

Show that $p \to \Box p$ is invalid (and thus $p \not\models \Box p$), by giving a countermodel. Explain why this doesn't contradict the previous exercise.

Exercise 2.8

Show that for any sentences A, B, if $\models A \to B$, then also $\models \Box A \to \Box B$.

Earlier in this section, I showed that $\Box p \to p$ and $\Box p \to \Box\Box p$ are valid. The arguments I gave easily generalise to other sentences in place of p. All instances of $\Box A \to A$ and $\Box A \to \Box\Box A$ are valid.

You may remember these schemas from section 1.5. There I defined the system S5 by stipulating that it contains all instances of the following schemas:

(Dual)	$\neg\Diamond A \leftrightarrow \Box\neg A$
(T)	$\Box A \to A$
(K)	$\Box(A \to B) \to (\Box A \to \Box B)$
(4)	$\Box A \to \Box\Box A$
(5)	$\Diamond A \to \Box\Diamond A$

All instances of these schemas are valid by the definitions of the present chapter.

I also specified two "rules" for S5. The first says that any truth-functional consequence of any sentences in S5 is itself in S5. The second says that whenever a sentence A is in S5, then so is $\Box A$. As we will show in chapter 4, these rules preserve validity (as defined in the previous section). Indeed, we will show that the sentences

that come out as valid by our present definitions are precisely the sentences in S5.

You may pause a moment to ponder how this could be shown. In the meantime, let's prove a simpler fact to which I have appealed above (as well as on page 15 in the previous chapter): that replacing logically equivalent sentences inside a larger sentence never affects the larger sentence's truth-value at any world.

Observation 2.1: If A is an $\mathfrak{L}_M$-sentence and A' results from A by replacing a subsentence of A with a logically equivalent sentence, then A and A' are logically equivalent.

Proof. Remember that two sentences are logically equivalent if each entails the other. By definition 2.4, this means that the two sentences are true at the same worlds in every model.

Now let A be an arbitrary $\mathfrak{L}_M$-sentence and assume that A' results from A by replacing a subsentence of A with a logically equivalent sentence. To show that A and A' are equivalent, we first show that this holds for the special case where A is a sentence letter. Then we consider different ways in which A might be built up from simpler sentences and show that *if the observation holds for those simpler sentences*, then it also holds for A itself.

So assume that A is a sentence letter. In that case, A has no sentences as proper parts. The observation is vacuously true. (There is no way of turning p into a non-equivalent sentence by replacing a subsentence within p.)

Next we assume that A is a complex sentence and that the observation holds for all simpler sentences.

To begin, assume that A is the negation of another sentence B. So A is $\neg B$ and A' is $\neg B'$ for some sentence B' that is either equivalent to B (if B is the subsentence of A that has been replaced to yield A') or that results from B by replacing a subsentence within B by an equivalent sentence (if the subsentence of A that has been replaced to yield A' isn't B). In the latter case, our assumption that the observation holds for sentences simpler than A implies that B and B' are equivalent. So either way, B and B' are logically equivalent. They are true at the same worlds in every model. By clause (b) of definition 2.2, it follows that A and A' are also true at the same worlds in every model.

Essentially the same reasoning applies in the case where A is a conjunction

$B \wedge C$, a disjunction $B \vee C$, a conditional $B \rightarrow C$, a biconditional $B \leftrightarrow C$, a box sentence $\Box B$, and a diamond sentence $\Diamond B$. I won't bore you by going through all of them. Here is the case for $\Box B$.

Assume that A has the form $\Box B$. So A is $\Box B$ and A' is $\Box B'$ for some sentence B' that is equivalent to B (by the same reasoning as before). By clause (g) of definition 2.2 it follows that A and A' are also equivalent. □

The style of proof I have used here is called an **induction on complexity**. It is widely used when reasoning about formal languages. In general, if you want to show that every sentence of a language has some property, it suffices to show that (i) all atomic sentences in the language have the property, and (ii) *if* all sentences that are simpler than a certain sentence have the property then so does that sentence. (In this context, the assumption that all simpler sentences have the property is called the *induction hypothesis*.)

2.5 Trees

I will now introduce a streamlined method for working through definition 2.2 to check whether a sentence is valid: the method of **analytic tableau** or **tree proofs**. (You may be familiar with this method for non-modal logic. If so, good. If not, no problem.)

The tree method is in the first place a method for finding countermodels. It is best introduced by example.

Let's try to find a countermodel for $\Diamond p \rightarrow \Box p$. That is, we want to construct a model in which there is some world w at which $\Diamond p \rightarrow \Box p$ is false. We start our search by assuming that the *negation* of $\Diamond p \rightarrow \Box p$ is *true* at w. We write this down as follows.

1. $\neg(\Diamond p \rightarrow \Box p)$ (w) (Ass.)

'1.' and '(Ass.)' are for book-keeping; 'Ass.' is short for 'Assumption', since we're *assuming* that $\neg(\Diamond p \rightarrow \Box p)$ is true at w. Now we unfold this assumption in accordance with definition 2.2. The definition tells us that a conditional $A \rightarrow B$ is false at a world w iff the antecedent A is true at w and the consequent B is false at w. So the assumption on line 1 implies that $\Diamond p$ is true at w and that $\Box p$ is false at w. We expand our "tree" (or "tableau") by adding these consequences.

1.	$\neg(\Diamond p \rightarrow \Box p)$	(w)	(Ass.)	✓
2.	$\Diamond p$	(w)	(1)	
3.	$\neg\Box p$	(w)	(1)	

I have ticked off line 1 (with '✓') to mark that we won't need to look at it again. All the information in line 1 is contained in lines 2 and 3. The parenthetical '(1)' at lines 2 and 3 reminds us that these assumptions are derived from line 1.

We continue drawing out further consequences. What does the truth of $\Diamond p$ at w imply for the subsentence p? By definition 2.2, there must be some world – let's call it v – at which p is true.

1.	$\neg(\Diamond p \rightarrow \Box p)$	(w)	(Ass.)	✓
2.	$\Diamond p$	(w)	(1)	✓
3.	$\neg\Box p$	(w)	(1)	
4.	p	(v)	(2)	

Line 3 claims that $\Box p$ is false at w. By definition 2.2, $\Box p$ is true at w iff p is true at all worlds. So if $\Box p$ is false at w, there must be some world at which p is false. Let's introduce such a world, naming it u. Our tree looks as follows.

1.	$\neg(\Diamond p \rightarrow \Box p)$	(w)	(Ass.)	✓
2.	$\Diamond p$	(w)	(1)	✓
3.	$\neg\Box p$	(w)	(1)	✓
4.	p	(v)	(2)	
5.	$\neg p$	(u)	(3)	

Now the only unprocessed lines are assumptions about sentence letters and negations of sentence letters. Sentence letters don't have (non-trivial) subsentences, so we can't use definition 2.2 to further break down 4 or 5. The tree is complete. We have found a countermodel for $\Diamond p \rightarrow \Box p$.

Let's read off the countermodel. There are three worlds in our tree: w, v, and u. So $W = \{w, u, v\}$. By line 4, p is true at v. By line 5, p is false at u. We don't know whether p is true or false at w, and it doesn't matter – otherwise the tree would say. Let's say that $V(p) = \{v\}$. As you can verify, $\Diamond p \rightarrow \Box p$ is indeed false at world w in this model.

One more example, before I state the general rules. Let's try to find a countermodel for $\Box(p \to q) \to (p \to \Box q)$. That's another conditional, so we begin as before.

1.	$\neg(\Box(p \to q) \to (p \to \Box q))$	(w)	(Ass.) ✓
2.	$\Box(p \to q)$	(w)	(1)
3.	$\neg(p \to \Box q)$	(w)	(1)

Line 1 assumes that the negation of the conditional is true at some world w. Lines 2 and 3 break down this assumption, using the fact that $\neg(A \to B)$ is true (at a world) iff A is true and B false. We could deal with line 2 next, but it's better to ignore it for the moment and process 3 first, which is yet another negated conditional.

4.	p	(w)	(3)
5.	$\neg\Box q$	(w)	(3)

Line 5 tells us that $\Box q$ is false at w. We can infer that there is a world – call it v – at which q is false.

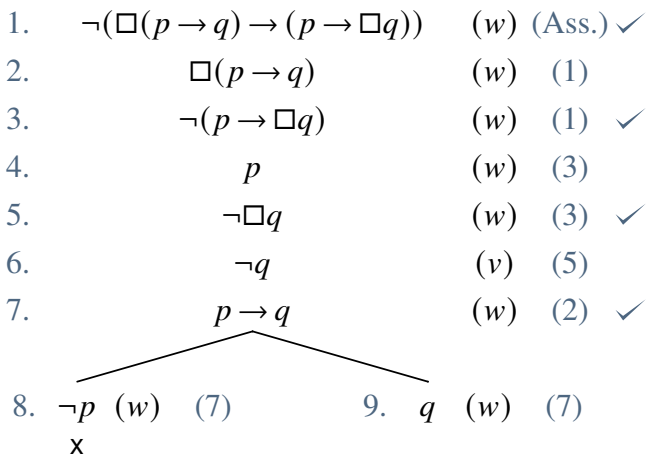
6.	$\neg q$	(v)	(5)

Now we need to return to line 2. What can we infer from the hypothesis that $\Box(p \to q)$ is true at w about the subsentence $p \to q$? By definition 2.2, $p \to q$ must be true at *every* world. So, in particular, $p \to q$ must be true at w. Let's write that down. We'll add another line for v later, so we don't check off node 2.

7.	$p \to q$	(w)	(2)

If you are used to proofs in the natural deduction style, you may now be tempted to apply Modus Ponens and infer that q is true at w, from lines 4 and 7. In the tree method, however, we try not to draw inferences from multiple premises. We simply look at any lines that can still be processed and check what definition 2.2 tells us about the immediate subsentences of the sentence on that line. So we process line 7 without looking at line 4.

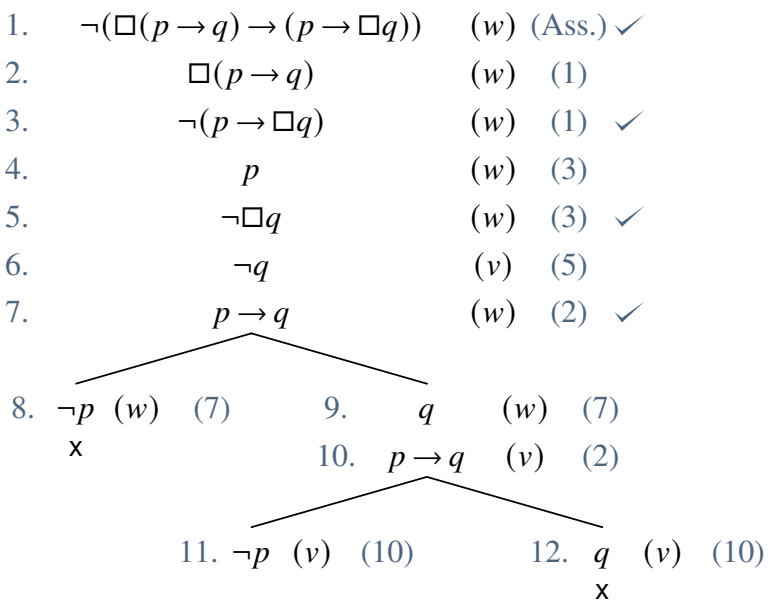
What can we infer from the truth of $p \to q$ at w about the subsentences p and q? By definition 2.2, $p \to q$ is true at w if *either* p is false at w *or* q is true at w. We have to keep track of both possibilities. So our (upside down) tree will branch. Here is the full tree at its present stage.



So far, I have called the numbered items on a tree 'lines'. The proper term is **nodes**. Since nodes 8 and 9 are visually on the same line, it would be confusing to call them lines. While we're at it, a **branch** of a tree is series of nodes that extends from the top (or "root") node all the way down to a node below which there is no other node. The present tree has two branches, both of which contain 8 nodes.

What does this tree tell us? Remember that our aim is to construct a model in which the sentence at node 1 is true at world w. So far, the tree tells us that there are two worlds w and v in this model; nodes 4 and 5 tell us something about the interpretation function in the model: p is true at w, q is false at v. After node 7, the tree branches, meaning that there are two ways of extending the model we have construed so far. On the left branch, we assume that p is false at w. On the right branch, we assume that q is true at w. But hold on. We already know that p is true at w (from node 4). There's no model in which p is both true and false at w. So the possibility explored on the left branch is a dead-end. it doesn't lead to a countermodel. That's why I've *closed* the left branch by drawing a cross below node 8.

We continue on the right-hand branch. Here we expand node 2 again, this time for world v, which leads to another branching.



On the right-most branch, q is true at v (by node 12) but also false at v (by node 6), so that branch is closed. But the middle possibility is still open, and there are no more assumptions to unfold. We have found a countermodel.

The countermodel is given by all the assumptions *on the middle branch*, the one that remained open. (The other branches were dead-ends and can be ignored.) We have two worlds, $W = \{w, v\}$. The interpretation function V makes p true at w (node 4) and false at v (node 11); q is also true at w (node 9) and false at v (node). Again, you may verify that the sentence on node 1 is true at world w in this model.

Now for the general rules.

In order to find a countermodel for a sentence A with the help of the tree method, you always begin by assuming that the *negation* of A is true at world w:

1. $\neg A$ (w) (Ass.)

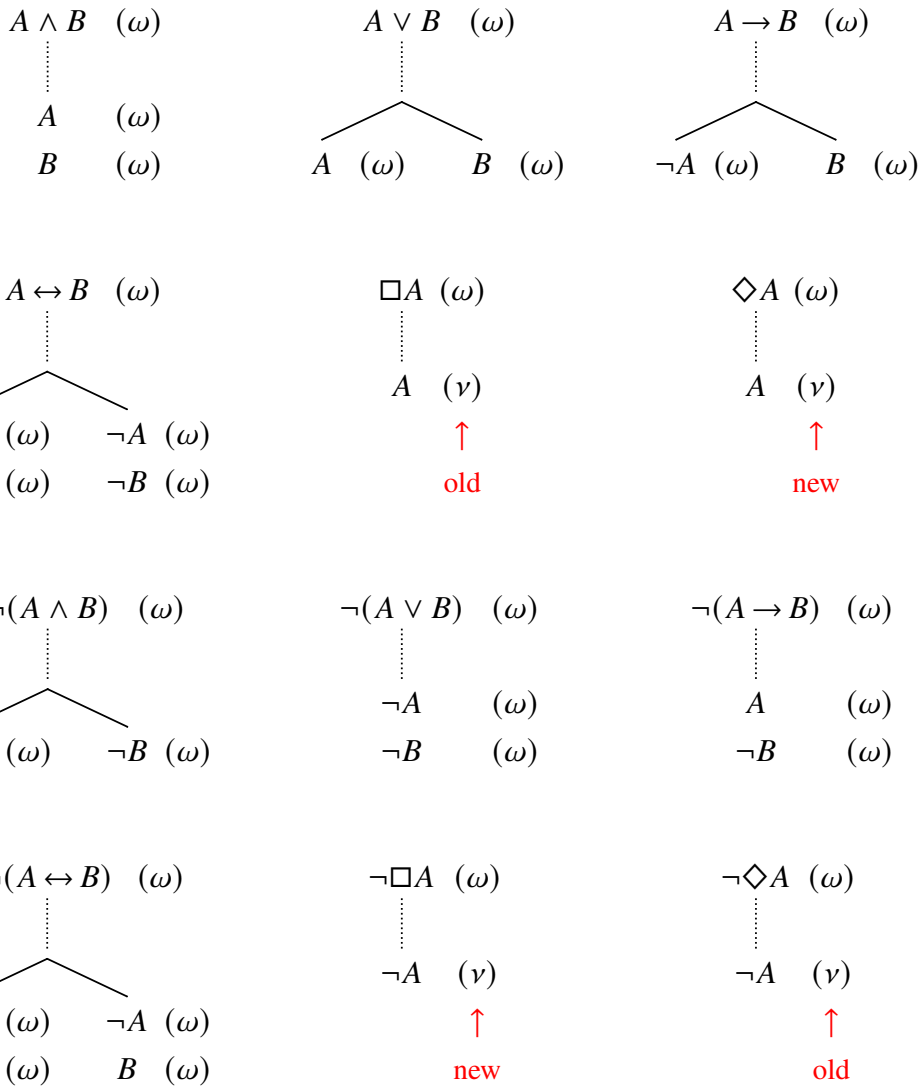
You then expand this node, and every new node that appears on the tree, until no more nodes can be expanded.

To expand a node with a non-negated sentence, you consider what the truth of the sentence at the node's world implies for the truth-value of the sentence's immediate parts. The result may be added to the end of any open branch containing the node.

(The immediate parts of a sentence of the form $A \wedge B$, $A \vee B$, $A \to B$, or $A \leftrightarrow B$ are the corresponding sentences A and B; the only immediate part of $\Box A$, $\Diamond A$, and $\neg A$ is A.)

To expand a node with a negation $\neg A$, you consider what the falsity of the relevant sentence A at the node's world implies for the immediate parts of A. The result may again be added to the end of any open branch containing the node.

The following diagrams summarize how the different kinds of nodes are expanded.



$$\begin{array}{ll} \neg\neg A & (\omega) \\ \vdots & \\ A & (\omega) \end{array}$$

If a branch of a tree contains a sentence A as well as its negation $\neg A$, for the same world ω, then the branch is *closed* with an x at the bottom.

The rule for $\Box A$ says that from the assumption that $\Box A$ is true at a world ω, you may infer that A is true at any world *v that already occurs on the branch to which the new node is added*. So you're not allowed to introduce a new world variable ('v', 'u', etc.) when expanding $\Box A$ nodes. The same is true for $\neg\Diamond A$ nodes (which by duality means the same as $\Box\neg A$). When you expand a $\Diamond A$ node (or a $\neg\Box A$ node), by contrast, you must introduce a new world variable.

Nodes of type $\Box A$ and $\neg\Diamond A$ can be expanded several times, once for every world variable on any branch containing the node.

If you have expanded a node that is not of type $\Box A$ or $\neg\Diamond A$, and you have added the new nodes to every open branch containing the node, then you can tick off the node. You don't need to look at it again. Nodes of type $\Box A$ and $\neg\Diamond A$ nodes are never ticked off.

If no more rules can be applied, the tree is complete. Any open branch on a complete tree defines a countermodel for the target sentence.

Exercise 2.9

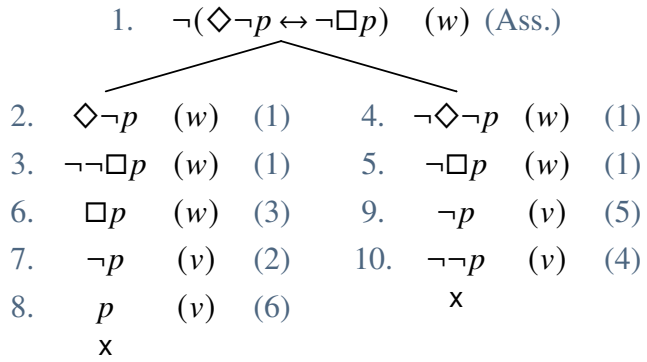
Use the tree method to find countermodels for the following sentences:

(a) $p \to \Box(p \vee q)$
(b) $\Box p \vee \Box\neg p$
(c) $\Diamond(p \to q) \to (\Diamond p \to \Diamond q)$
(d) $p \to q$
(e) $\Box\Diamond p \to p$

What if all branches on a tree close? Then there is no countermodel for the target sentence. If there is no countermodel for a sentence, then the sentence is valid. This is how the tree method is used to show that sentences are valid.

The following tree shows that $\Diamond\neg p \leftrightarrow \neg\Box p$ is valid. Make sure you understand each step. (I've omitted the check marks since these are only useful during the

construction phase.)



A similar tree could obviously be drawn for $\Diamond\neg q \leftrightarrow \neg\Box q$, and for any other formula of the form $\Diamond\neg A \leftrightarrow \neg\Box A$: we would simply replace each occurrence of p on the tree with A.

To show that all instances of a schema are valid, we can also directly draw **schematic trees** in which we use schematic variables 'A', 'B', 'C' instead of sentence letters.

Exercise 2.10

Use the tree method to show that all instances of the following schemas are valid.

(K) $\Box(A \to B) \to (\Box A \to \Box B)$
(T) $\Box A \to A$
(D) $\Box A \to \Diamond A$
(4) $\Box A \to \Box\Box A$
(5) $\Diamond A \to \Box\Diamond A$
(G) $\Diamond\Box A \to \Box\Diamond A$

Exercise 2.11

For each of the following sentences, either show that it is valid or give a countermodel to show that it is invalid:

(a) $p \to \Box\Diamond p$
(b) $\Diamond\Diamond p \to \Diamond p$
(c) $\Diamond(p \wedge q) \to (\Diamond p \wedge \Diamond q)$

(d) $(\Diamond p \wedge \Diamond q) \rightarrow \Diamond(p \wedge q)$
(e) $\Diamond(p \vee q) \leftrightarrow (\Diamond p \vee \Diamond q)$
(f) $\Box\Diamond p \rightarrow \Diamond\Box p$
(g) $(\Diamond p \rightarrow \Box q) \rightarrow (\Box p \rightarrow \Box q)$

When constructing a tree, you often have a choice of which node to expand next. In that case, a good idea is to start with any $\Diamond A$ or $\neg\Box A$ nodes. If there are none, choose a node of type $A \wedge B$, $\neg(A \vee B)$ or $\neg(A \rightarrow B)$. Choose a node of another type only if none of the above are available. This heuristic often helps to keep trees small, but it is not part of the official tree rules.

Exercise 2.12

Can we use the tree method to show that some premises $A_1, \ldots, A_n$ entail a conclusion B? Can we use it to show that two sentences A and B are equivalent?