Honors Logic, Lecture 10 - Modal Logic

Brian Weatherson

2022-09-30

What Modal Logic Is

The logics of **must** and **might**.

- Why plural? Because we do not assume that these words have a single determinate meaning.

Examples of Must

1. If $x = 2 + 2$, then x must equal 4.
2. If something is a cat, then it must be a mammal.
3. If the gardener is innocent, then it must be the butler who did it.
4. You must drive under 70mph on I-94.
5. You must keep your promises.
6. If you set out a knife and fork, the fork must go on the left.

- To my ears, 1 is **logical** necessity, 2 is **metaphysical** necessity, 3 is **epistemic** necessity, 4 is **legal** necessity, 5 is **moral** (or **deontic**) necessity and 6 is **etiquette** necessity.

Examples of May/Might

1. If x is prime, then x might be even.
2. If x is a cat, then x might be male.
3. It might be the butler or the gardener that did it.
4. You may drive at any speed below 30mph on State Street.
5. You may lie to save a friend's life.
6. You may use white napkins or red napkins.

- To my ears, 1 is **logical** possibility, 2 is **metaphysical** possibility, 3 is **epistemic** possibility, 4 is **legal** possibility, 5 is **moral** (or **deontic**) possibility and 6 is **etiquette** possibility (though I'm not sure about any of these).

Logics

Consider this very general claim.

If something must be true, then it is true.

- That's true on the logical, epistemic and metaphysical interpretations of modality. Indeed, it's something like a logical truth of those domains.
- But it is very much not true on the legal, moral or etiquette interpretations.

So we want some logics where it is a logical truth, and some where it is not.

Language

We extend our language with two new operators: $\Box$ and $\Diamond$.

- If p is a sentence, so is $\Box p$ and so is $\Diamond p$.
- These mean, respectively, that p must be true, and that p might be true.
- We interpret these somewhat similar to negations; they just bind what they are immediately next to.
- So $\Box p \to q$ means $(\Box p) \to q$, not $\Box(p \to q)$.

Truth

What does it take for these sentences to be true?

Worlds

We start with Leibniz's idea that necessity is truth in all possible worlds.

- Leibniz was interested in metaphysical necessity, so we'll have to qualify this a little, but it's a good idea.
- So instead of saying that each proposition simply has a truth value, we'll say that there are many **worlds**, and at each world each proposition has a truth value.
- But don't assume that propositions have the same truth value at each world.
- In one world I might be standing, and in another world I might be sitting.

What Are Worlds

We are well and truly not going to get into the metaphysics of worlds here.

- Indeed, they need not even be anything like possible worlds in the sense that metaphysicians usually care about.
- They might, for instance, be different times.
- All we care about is that they are things at which propositions can be true or false.

Valuations

A valuation function tells us which worlds atomic sentences are true at.

- These can be completely arbitrary; we don't put any restrictions on them.

Truth at a World

We want more generally a function that tells us whether a sentence is true at a particular world.

- For sentences built up using $\wedge, \vee, \rightarrow, \neg$, this is relatively easy.
- We just keep on using truth tables.
- So if at world w, A is true and B is false, then $A \wedge B$ is false and $A \vee B$ is true.

Modal Values

We also need values for these sentences:

- $\Box A$
- $\Diamond A$

It turns out these are more complicated - but not much more complicated.

Accessibility

The last part of our model is an **accessibility** relation between worlds.

- Again, this can be completely arbitrary.
- We don't yet put any restrictions on it.
- Notably, we don't assume that it is **reflexive**, **symmetric** or **transitive**

Properties of Relations

- R is reflexive iff for all x, xRx.
- R is symmetric iff for all x, y, if xRy then yRx.
- R is transitive iff for all x, y, z if xRy and yRz then xRz.

A lot of relations we care about have one or more of these properties, but not all do. Consider, for example, **admires** as an example of a relation with none of them.

Truth of Modal Formulas

A sentence $\Box A$ is true at a world x just in case the following condition is met:

- For all worlds y such that xRy, A is true at world y.

A sentence $\Diamond A$ is true at a world x just in case the following condition is met:

- For some world y such that xRy, A is true at world y.

Modal Truth

- Something is necessarily true iff it is true everywhere that is accessible.
- Something is possibly true iff it is true somewhere accessible.

We get back the Leibnizian idea that necessity is truth in all possible worlds if we assume the accessibility relation is the universal relation, i.e., xRy for all x, y.

Metaphysical Necessity

On this Leibnizian model, where all worlds can access all worlds, iterated modalities are rather uninteresting. These three sentences are true in the same worlds/models.

1. $\Box A$
2. $\Box\Box A$
3. $\Diamond\Box A$

- That's because if $\Box A$ is true at any world, then it is true at all worlds. In the general case, where we do not assume that R is universal, these are not equivalent.

Duality

These two claims are equivalent.

1. $\Box A$
2. $\neg\Diamond\neg A$

From 1 to 2: If $\Box A$ is true at x, then A is true for all y such that xRy. That means there is no y such that xRy and A is not true. That means there is no y such that xRy and $\neg A$ is true. That means $\Diamond\neg A$ is not true at w. That means $\neg\Diamond\neg A$ is true at x.

Duality

These two claims are equivalent.

1. $\Box A$
2. $\neg\Diamond\neg A$

From 2 to 1: If $\neg\Diamond\neg A$ is true at x, then $\Diamond\neg A$ is not true at w. So there is no world y such that xRy and $\neg A$ is true at y. So at all worlds y such that xRy, $\neg A$ is not true. So at all worlds y such that xRy, A is true. So $\Box A$ is true at x.

Duality

These two claims are also equivalent, though I will not prove this.

1. $\Diamond A$
2. $\neg\Box\neg A$

Big Picture

These claims are both logically true.

1. $\Box\neg A \leftrightarrow \neg\Diamond A$
2. $\Diamond\neg A \leftrightarrow \neg\Box A$

To move a negation sign outside of a modal operator, either $\Box$ or $\Diamond$, you have to rotate the operator by 45 degrees.

Normality

This sentence is also true no matter what the model looks like, and no matter what sentence A is.

$$\Box(A \rightarrow B) \rightarrow (\Box A \rightarrow \Box B)$$

- Assume it is false at w.
- So $\Box(A \rightarrow B)$ is true at w and $(\Box A \rightarrow \Box B)$ is false at w.
- So $\Box A$ is true at w and $\Box B$ is false at w.
- So at every where y such that wRy, A must be true (since $\Box A$ is true at w), and $A \rightarrow B$ must be true (since $\Box(A \rightarrow B)$ is true at w).
- If A and $A \rightarrow B$ are true at y, B must be true at y as well.
- But this implies that B is true all y such that wRy, contradicting the assumption that $\Box B$ is false at w.

Normality

This principle has a very important role in the history of modal logics.

$$\Box(A \rightarrow B) \rightarrow (\Box A \rightarrow \Box B)$$

Having this be an axiom is one of two conditions on what have come to be called **normal** modal logics.

Models

Models have three parts:

1. A set of worlds, typically called W.
2. A binary accessibility relation on those worlds, typically called R.
3. A valuation function on those worlds, typically called V.

We'll write the models as $\langle W, R, V\rangle$.

Valuations

V is a function from atomic sentence letters to subsets of W.

- It tells you when the atomic sentences are true.
- When an atomic sentence is not true, it is false.

Truth at a Point

The general theory of truth is built up in stages from the basic theory. Assume we have a model $\langle W, R, V \rangle$, and a point $w \in W$, and are asking whether an arbitrary sentence is true at w in $\langle W, R, V \rangle$.

- p is true at w iff $w \in V(p)$.
- $\neg A$ is true at w iff A is not true at w.
- $A \wedge B$ is true at w iff A is true and w and B is true at w.
- $A \vee B$ is true at w iff A is true and w or B is true at w.
- $A \rightarrow B$ is true at w iff A is false at w or B is true at w.

This just leaves the modal formulae. I'll set out the rules, then do some worked examples.

Necessary Truth at a Point

First we'll do $\Box A$.

- I'll read this as 'Box A'.
- Intuitively, it means **It must be that A**, where **must** could be a metaphysical necessity, or an epistemic necessity, or a moral necessity, or anything else.
- And it is true at w just in case A is true at every world y such that wRy.
- Necessary truth is truth at all accessible worlds.

Possible Truth at a Point

Now we'll do $\Diamond A$.

- I'll read this as 'Diamond A'.
- Intuitively, it means **It might be that A**, where **might** could be a metaphysical necessity, or an epistemic necessity, or a moral necessity, or anything else.
- And it is true at w just in case A is true at some world y such that wRy.
- Possible truth is truth at some accessible world.

Iterated Modalities

We can run these rules in sequence.

- What does it take for $\Box\Box A$ to be true at w?
- It is for $\Box A$ to be true at every y such that wRy.
- And that means that A has to be true at every world z such that yRz (for any y such that wRy).

Access

We can think, a little picturesquely, as the accessibility relation being a 'step' between worlds.

- If wRy, then you can 'step' from w to y.
- $\Box A$ means that anywhere you can step to from w is a world where A is true.
- And $\Box\Box A$ means that anywhere you can get to in two steps from w is a world where A is true.

Iterated Modalities

We can run the rules in sequence.

- What does it take for $\Diamond\Diamond A$ to be true at w?
- It is for $\Diamond A$ to be true at some y such that wRy.
- And that means that A has to be true at some world z such that yRz (for some y such that wRy).
- In the picturesque terms, you can get from w to an A-world in two steps.

Mixed Modalities

What does it mean for $\Diamond\Box A$ to be true at w?

- There is some accessible world where $\Box A$ is true.
- That is, there is some accessible world such that everywhere you can go from there, A is true.

Mixed Modalities

What does it mean for $\Box\Diamond A$ to be true at w?

- At all accessible worlds, $\Diamond A$ is true.
- That is, wherever you go, you can get to there is some accessible world such that everywhere you can go from there, A is true.

Longer Sentences

What does it mean for $\Box(p \vee (q \to \Diamond r))$ to be true at w?

- It's for $p \vee (q \to \Diamond r)$ to be true everywhere you can get to (in one step) from w.
- That is, at every one of those worlds, either p is true or $q \to \Diamond r$ is true.
- That is, at every one of those worlds, either p is true, or q is false, or $\Diamond r$ is true.
- That is, at every one of those worlds, either p is true, or q is false, or there is some world you can get to where r is true.

Box and connectives

The general rule is just to apply the rules for sentences inside the brackets at each world in W, and then apply the rule for $\Box$ or $\Diamond$. But there are three special cases worth thinking about.

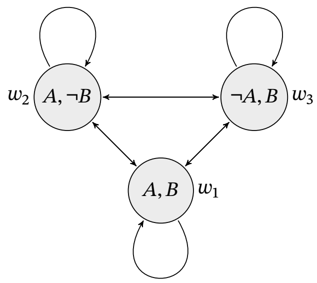
- $\Box(A \wedge B)$ means that all accessible worlds are A and B worlds.
- $\Box(A \vee B)$ means that all accessible worlds make at least one of A and B true.
- $\Box(A \to B)$ means that all accessible A-worlds are B-worlds.

We'll use that last one a lot.

I'm going to go through some sentences, and show how there can be models where each of them is false. Here is what we'll cover

1. $\Box(A \vee B) \to (\Box A \vee \Box B)$
2. $(\Diamond A \wedge \Diamond B) \to \Diamond(A \wedge B)$
3. $A \to \Box A$
4. $\Box A \to A$
5. $\Box\Diamond A \to B$
6. $\Box\Diamond A \to A$
7. $\Box\Box A \to \Box A$
8. $\Box A \to \Box\Box A$
9. $\Box\Diamond A \to \Diamond\Box A$
10. $\Box A \to \Diamond A$

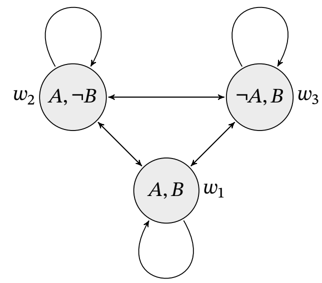
$\Box(A \vee B) \to (\Box A \vee \Box B)$



At all points, either A or B is true, so $\Box(A \vee B)$ is true.

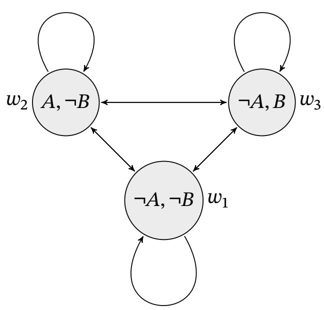
But $\Box A$ and $\Box B$ are false everywhere. So the conditional is false everywhere.

$\Box(A \vee B) \to (\Box A \vee \Box B)$



Note that this is overkill. We just need to show that the formula can be false somewhere in order to show that it is not a theorem.

$(\Diamond A \wedge \Diamond B) \to \Diamond(A \wedge B)$

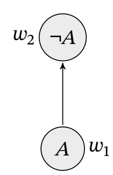


At w_1, we have $\Diamond A \wedge \Diamond B$ true.

But nowhere is $A \wedge B$ true, so $\Diamond(A \wedge B)$ is false at w_1. So the conditional is false.

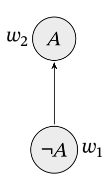
Again, this is overkill.

$A \to \Box A$



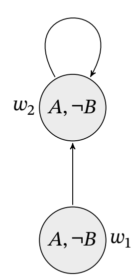
- At w_1 A is true.
- But $\Box A$ is false, since w_1 can access w_2, and A is false there.
- So $A \to \Box A$ is false.

$\Box A \to A$



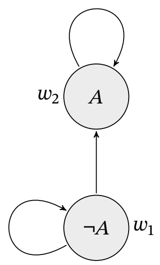
- At w_1 $\Box A$ is true. The only accessible world is w_2, and A is true there.
- But A is false there.
- So $\Box A \to A$ is false.

$\Box\Diamond A \to B$



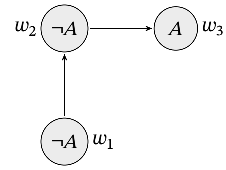
- At w_1 $\Box\Diamond A$ is true. The only accessible world is w_2, and $\Diamond A$ is true there. (Why?)
- But B is false at w_1.
- So $\Box\Diamond A \to B$ is false.

$\Box\Diamond A \to A$



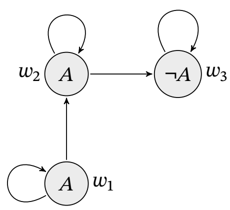
- At w_1 $\Box\Diamond A$ is true. At every world, w_2 is accessible, and A is true there.
- But A is false at w_1.
- So $\Box\Diamond A \to A$ is false at w_1.

$\Box\Box A \to \Box A$



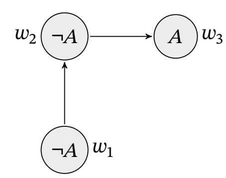
- The only world w_2 can access is w_3, and A is true there, so $\Box A$ is true at w_2.
- The only world w_1 can access is w_2, and $\Box A$ is true there, so $\Box\Box A$ is true at w_1.
- But $\Box A$ is false at w_1.
- So $\Box\Box A \to \Box A$ is false at w_1.

$\Box A \to \Box\Box A$



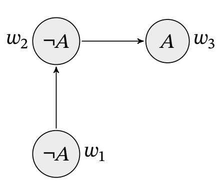
- Since A is false at w_3, and w_2 can access w_3, $\Box A$ is false at w_2.
- Since $\Box A$ is false at w_2, and w_1 can access w_2, $\Box\Box A$ is false at w_1.
- But $\Box A$ is true at w_1.
- So $\Box A \to \Box\Box A$ is false at w_1.

$\Box A \to \Diamond A$



- Focus on w_3.
- There is no accessible world where A is false, so $\Box A$ is true there.
- But there is no accessible world where A is true, so $\Diamond A$ is false there.
- So $\Box A \to \Diamond A$ is false there.

$\Box A \to \Diamond A$



Whenever there are no accessible worlds, the following two weird things happen.

1. All $\Box$-sentences (i.e., sentences that start with a $\Box$ that takes scope over the whole sentence) are true.
2. All $\Diamond$-sentences (i.e., sentences that start with a $\Diamond$ that takes scope over the whole sentence) are false.