Modal logic

Michael Franke

Notions covered: language of modal logic, (pointed) modal models, truth in pointed models, belief and knowledge models.

Propositional logic and predicate logic give us formulas to talk about the *here and now*. For example, the sentence:

$$\underbrace{\text{The earth is round}}_{p}$$

is to be understood as a claim about how the world is like. If we wanted to check whether this sentence is true or false, we would have to go search *this* world, the *actual world* that we live in.

Yet, much of human language involves imaginary circumstances, ways in which the world could have been but isn't. For example, in order to evaluate whether the following sentences are true:

$$\underbrace{\text{It is logically necessary that}}_{\Box}\ \underbrace{\text{the earth is round}}_{p}$$

$$\underbrace{\text{Alex believes that}}_{\Box_a}\ \underbrace{\text{the earth is round}}_{p}$$

it does not suffice to just check the actual shape of the earth. We need to consider different (imaginary) possibilities of how the world could have been. We call these ways the world could have been *possible worlds*. To determine the truth or falsity of sentences like the above, we need to determine whether certain possible worlds (here: worlds in which the earth is not round) have certain properties, such as whether they are compatible with the rules of logic or with what Alex considers possible.

Many natural language expressions carry a modal meaning component. In English, there are words like *should*, *could*, *possibly*, *need to*, *be required* and many more. In order to capture such modal meaning, modal logics provide a formal language with which we can express what is true in the actual world and what is true in other possible worlds. To do so, modal logics usually have (at least) these two *modal operators*:

operator	neutral paraphrase	natural language cues
$\Diamond$	It is possible that ...	*might*, *may*, *conceivably*, ...
$\Box$	It is necessary that ...	*must*, *have to*, *necessarily*, ...

There are different kinds of modal logic, depending on the kind of modal meaning we would like to capture. Indeed, there are many different kinds of *modalities*, also referred to as *modal flavors*:

(i) what one ought to do given one's parents' rules and regulations;

(ii) what one ought to do given a country's law;

(iii) what is logically necessary;

(iv) what an agent believes to be true or possible;

... and more.

Different types of modality require different assumptions about the logical properties of the modal operators. In this way, we can think of modal logic as a tool for *conceptual analysis* of the logical properties that characterize concepts like *belief*, *logical necessity*, *obligation* etc. Here, we will have a look at a modal propositional logic for talking about beliefs and knowledge, a so-called *epistemic modal logic*.

1 The language of epistemic modal logic

Modal propositional logic (ModLog) extends propositional logic by including the new operators $\Diamond$ and $\Box$. To make things more interesting, we will have one pair of operators $\Diamond_i$ and $\Box_i$ where i is a variable for one of several *agents*. We will interpret a formula $\Diamond_i\varphi$ as: "agent i considers φ to be possible;" and $\Box_i\varphi$ as: "agent i believes that φ is true." In this way, we will be able to formally express sentences like "Alex believes that Bo considers it possible that the earth is not round" as $\Box_a\Diamond_b\neg p$.

1.1 Formulas

Let $\mathcal{P}$ be a set of proposition letters and let $\mathcal{A}$ be a set of agents. The language $\mathfrak{L}_{\mathcal{P},\mathcal{A}}$ of ModLog is the set of all *formulas* which are recursively defined as follows:

(i) Every proposition letter is a formula.

(ii) If φ is a formula, so is $\neg\varphi$, $\Box_i\varphi$ and $\Diamond_i\varphi$ for each $i \in \mathcal{A}$.

(iii) If φ and ψ are formulas, so are:
a. $(\varphi \wedge \psi)$ b. $(\varphi \vee \psi)$ c. $(\varphi \to \psi)$ d. $(\varphi \leftrightarrow \psi)$

(iv) Anything that cannot be constructed by (i)–(iii) is not a formula.

Here are examples of well-formed formulas of our multi-agent epistemic logic with paraphrases (where agent a is Alex, and b is Bo):

$\Box_a(p \to q)$	Alex believes that if p, then q.
$p \to \Diamond_a\Box_b q$	If p is true, then Alex considers it possible that Bo believes that q.
$\Box_a p \wedge \Box_b p \wedge \neg\Box_a\Box_b p$	Both Alex and Bo believe that p, but Alex doesn't believe that Bo believes it.

1.2 *Semantics of modal logic*

The semantics of ModLog is defined in terms of so-called *modal models*. Similar to PropLog and PredLog, a modal model will tell us whether any given formula is true or false. But while models in PropLog and PredLog were representations of just one possible world, modal models comprise (possibly: infinitely) many possible worlds all at once.

Modal models. A *modal model* for the language $\mathfrak{L}_{\mathcal{P},\mathcal{A}}$ is a triple $\mathcal{M} = \langle W, V, (R_i)_{i \in \mathcal{A}} \rangle$ such that:[1]

W is a set of *(possible) worlds*,

$V : W \times \mathcal{P} \to \{0, 1\}$ is a *valuation function* assigning a unique truth value to every proposition letter for every possible world, and

for each agent $i \in \mathcal{A}$, $R_i \subseteq W \times W$ is an *accessibility relation* between possible worlds.

Accessibility relations are interpreted as follows: if the actual world is w, then agent i considers possible all the worlds that can be "accessed" or "seen" via relation R_i from w. We write wR_iv for $\langle w, v \rangle \in R_i$ and introduce the notation:

$$R_i(w) = \{v \in W \mid wR_iv\}$$

to refer to the set of *accessible worlds* for agent i from world w.

Single-agent example. For a simple example, let us just consider a case with a single agent first, which is Alex: $\mathcal{A} = \{a\}$. We are interested in Alex's beliefs about tomorrow's weather. For simplicity, we consider three possible states of the weather: it's either rainy, cloudy or sunny. We can model this in terms of three proposition letters: $\mathcal{P} = \{r, c, s\}$. For simplicity, let's consider just four possible worlds, $W = \{w_1, w_2, w_3, w_4\}$. We need to define a valuation function that maps every pair consisting of a possible world and a proposition letter onto a truth value. This could be written like so: $V(w_1, r) = 1$, $V(w_1, c) = 0$ and so on. Here is more compact representation of the valuation function we will use:[2]

world	r	c	s
w_1	1	0	0
w_2	0	1	0
w_3	0	0	1
w_4	1	0	0

Let's furthermore assume the following accessibility relation:

$$R_a = \{\langle w_1, w_2 \rangle, \langle w_1, w_3 \rangle, \langle w_2, w_2 \rangle, \langle w_2, w_3 \rangle, \langle w_3, w_2 \rangle, \langle w_3, w_3 \rangle, \langle w_4, w_4 \rangle\}$$

[1] Modal models are also often called *Kripke structures*, named after the logician and philosopher of language Saul Kripke.

[2] This looks suspiciously like a truth-table, but it is not. Not all logically possible worlds are listed (because we do not need them in this example). Also, there are rows (w_1 and w_4) with exactly the same truth-value assignments (we need them to model different beliefs, see below).

Figure 1 represents the whole modal model in a much more intelligible way. The diagram shows, for each possible world, the set of all proposition letters that are true in this world. Labelled arrows are used to represent the accessibility relation for each agent. Since we only consider a single agent, all arrows are (superfluously) labelled with a.

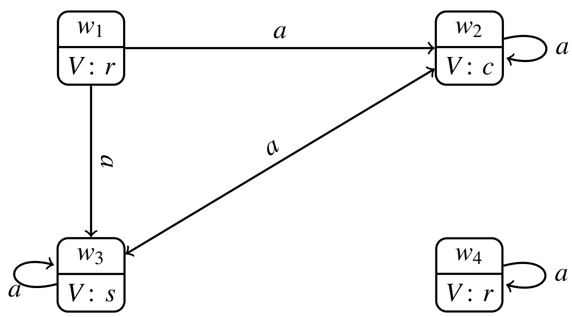

Figure 1: Example of a modal model for Alex's beliefs about tomorrow's weather. In world w_1 Alex (falsely) believes that it will not rain. In world w_4, Alex believes (correctly) that it will rain.

What beliefs does Alex hold according to the model in Figure 1? — Actually, that depends on which of the four worlds we consider to be the *reference world*, i.e., the world we consider to be actual for the purpose of analysis. If w_1 is the reference world (we can also say: "in world w_1"), Alex believes that it might be cloudy and that it might be sunny, but that it will not rain. This is because the worlds accessible from w_1, according to Alex's accessibility relation, include only the two worlds w_2 and w_3 where it is cloudy and sunny respectively: $R_a(w_1) = \{w_2, w_3\}$. So, in world w_1 Alex *rules out* the possibility of rain (even though that is the true weather for tomorrow, unbeknownst to Alex); Alex *rules in* the possibilities of cloudy and sunny weather, so that Alex is, after all, not entirely sure about the weather.

In worlds w_2 and w_3, Alex holds the exact same beliefs as in w_1. This is because we have the same set of accessible worlds for the first three worlds:

$$R_a(w_1) = R_a(w_2) = R_a(w_3) = \{w_2, w_3\}$$

In w_1 these beliefs are false (w_1 is a rain-world, but Alex excludes rain), in w_2 and w_3 they are not (e.g., w_2 is a cloudy world and Alex considers cloudy weather to be possible).

Finally, in world w_4 Alex's accessible worlds are $R_a(w_4) = \{w_4\}$. So, in w_4, Alex entertains only one relevant possibility. Alex is maximally opinionated, i.e., not uncertain at all. Alex believes (truthfully) that it will rain.

Truth conditions for ModLog. Truth conditions for formulas of ModLog are defined relative to *pointed models*. If $\mathcal{M} = \langle W, V, (R_a)_{a \in \mathcal{A}} \rangle$ and $w \in W$, the pair $\mathcal{M}, w$ is a pointed model. A valuation function $V_{\mathcal{M},w}$ for a pointed model, assigns truth values to each formula of ModLog as follows: For proposition letters $p \in \mathcal{P}$, the model's internal valuation function V decides on truth or falsity:[3]

[3] Notice that the function $V_{\mathcal{M},w}$ on the left-hand side is the "global" valuation function that builds on the pointed model to assign truth values to *all* formulas, while the valuation function V on the right-hand side is the valuation function *inside* of model $\mathcal{M}$ gives truth values to *only* proposition letters (for a given world).

$$V_{\mathcal{M},w}(p) = 1 \text{ iff } V(w,p) = 1$$

Formulas with main operators from propositional logic are treated as before:

$$\begin{array}{ll}
V_{\mathcal{M},w}(\neg\varphi) = 1 & \text{iff } V_{\mathcal{M},w}(\varphi) = 0 \\
V_{\mathcal{M},w}(\varphi \wedge \psi) = 1 & \text{iff } V_{\mathcal{M},w}(\varphi) = 1 \text{ and } V_{\mathcal{M},w}(\psi) = 1 \\
V_{\mathcal{M},w}(\varphi \vee \psi) = 1 & \text{iff } V_{\mathcal{M},w}(\varphi) = 1 \text{ or } V_{\mathcal{M},w}(\psi) = 1 \\
V_{\mathcal{M},w}(\varphi \rightarrow \psi) = 0 & \text{iff } V_{\mathcal{M},w}(\varphi) = 1 \text{ and } V_{\mathcal{M},w}(\psi) = 0 \\
V_{\mathcal{M},w}(\varphi \leftrightarrow \psi) = 1 & \text{iff } V_{\mathcal{M},w}(\varphi) = V_{\mathcal{M},w}(\psi)
\end{array}$$

What is new is the treatment of modal operators. For these, the accessibility relations are important.[4]

[4] Essentially, modal operators are like quantifiers from predicate logic, but quantifying over "accessible worlds."

$$\begin{array}{ll}
V_{\mathcal{M},w}(\Box_i\varphi) = 1 & \text{iff } V_{\mathcal{M},v}(\varphi) = 1 \text{ for all } v \in R_i(w) \\
V_{\mathcal{M},w}(\Diamond_i\varphi) = 1 & \text{iff } V_{\mathcal{M},v}(\ phi) = 1 \text{ for some } v \in R_i(w)
\end{array}$$

Crucially, the semantics for modal operators shifts the reference world, so to speak, switching from pointed model $\mathcal{M}, w$ to $\mathcal{M}, v$ for some accessible world $v \in R_i(w)$. In this way, we can also give a meaning to nested modal operators, as the following example demonstrates.

A more complex example. Figure 2 shows an example of a modal model for two agents ($\mathcal{A} = \{a, b\}$, Alex and Bo) and just two propositions letters ($\mathcal{P} = \{p, q\}$). This model only has three possible worlds $W = \{w, u, v\}$. The diagram shows, the set of all proposition letters that are true in each world. Labelled arrows are used to represent the accessibility relation for each agent.

$\Box_b(\neg p \wedge \neg q)$

"Bo believes that p and q are false"

$\neg p \wedge \Box_a p$

"Alex falsely believes that p is true"

$\neg\Box_a q \wedge \neg\Box_a\neg q$

"Alex is uncertain about q"

$\neg\Box_b p \wedge \Box_a\Box_b p$

"Alex falsely believes that Bo believes p"

$\neg\Box_a\Box_b q \wedge \neg\Box_a\neg\Box_b q$

"Alex is uncertain whether Bo believes q"

$\Box_a((q \rightarrow \Box_b q) \wedge (\neg q \rightarrow \Box_b\neg q))$

"Alex believes that Bo knows whether q"

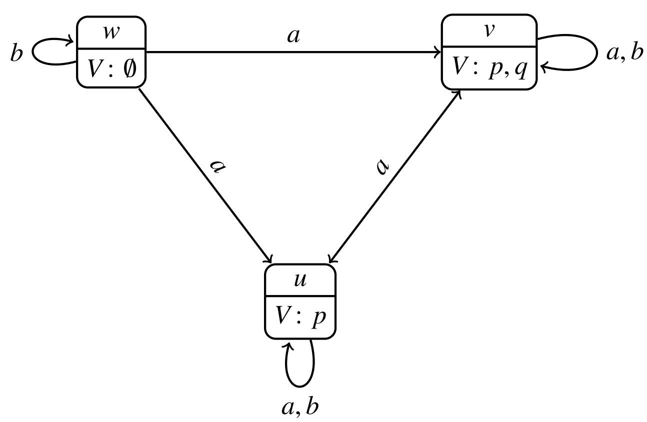

Figure 2: Example of a modal model with three worlds (w, v and u) for two agents (a and b) and two proposition letters (p and q).

The formulas on the left-hand side are all true in world w. To check that, for example, $V_{\mathcal{M},w}(\neg\Box_a\Box_b q \wedge \neg\Box_a\neg\Box_b q) = 1$, we can reason as follows:

$$\begin{aligned} & V_{\mathcal{M},w}(\neg\Box_a\Box_b q \wedge \neg\Box_a\neg\Box_b q) = 1 \\ \text{iff} \quad & V_{\mathcal{M},w}(\neg\Box_a\Box_b q) = 1 \quad \text{and} \quad V_{\mathcal{M},w}(\neg\Box_a\neg\Box_b q) = 1 \end{aligned}$$

Consider the first conjunct:

$$\begin{aligned} & V_{\mathcal{M},w}(\neg\Box_a\Box_b q) = 1 \\ \text{iff} \quad & \text{for some } w' \in R_a(w)\ V_{\mathcal{M},w'}(\Box_b q) = 0 \\ \text{iff} \quad & \text{for some } w' \in R_a(w') \text{ for some } w'' \in R_b(w')\ V_{\mathcal{M},w''}(q) = 1 \\ & \text{which is true if we set } w' = w'' = v \end{aligned}$$

Consider the second conjunct:

$$\begin{aligned} & V_{\mathcal{M},w}(\neg\Box_a\neg\Box_b q) = 1 \\ \text{iff} \quad & \text{for some } w' \in R_a(w)\ V_{\mathcal{M},w'}(\neg\Box_b q) = 0 \\ \text{iff} \quad & \text{for some } w' \in R_a(w)\ V_{\mathcal{M},w'}(\Box_b q) = 1 \\ \text{iff} \quad & \text{for some } w' \in R_a(w) \text{ for all } w'' \in R_b(w')\ V_{\mathcal{M},w'}(q) = 1 \\ & \text{which is true if } w' = u\text{, since } R_b(u) = \{u\} \text{ and } q \text{ is false in } u \end{aligned}$$

1.7 *Validity, entailment etc.*

The notions of validity, entailment, tautology, contradiction and contingency are defined essentially in the same way as for PropLog and PredLog. We only need to talk about truth at a pointed model. For example, we say that φ is a tautology of ModLog whenever $V_{\mathcal{M},w}(\varphi) = 1$ for all pointed models $\mathcal{M}, w$.

2 *The logic of rational belief*

So far, we have not put any constraints on the accessibility relations in our modal models. But, in order to represent beliefs that are "rational" in a crucial sense, we should. For example, we should rule out that $R_i(w) = \emptyset$ for some agent i and world w, because that would mean that agent i believes *anything* at this world, including logical contradictions (which isn't very rational, right?). This is to say that we would require rational agents to hold *consistent beliefs* that are contradiction-free. We can ensure contradiction-freedom by a constraint on admissible modal models, called *seriality*:

Seriality: for all i and w: $R_i(w) \neq \emptyset$

or, in other words, there is always at least one accessible world. We can then prove the following:

Claim 1. If a modal model's accessibility relation is serial, no agent can believe in a contradiction.

Proof. Let $\mathcal{M}, w$ be a pointed model with $R_i(w)$ the accessible worlds for agent i at w. Take an arbitrary contradiction φ. Being a contradiction entails that $V_{\mathfrak{M}'_{\mathcal{P},\mathcal{A}},v}(\varphi) = 0$ for all $\mathfrak{M}'_{\mathcal{P},\mathcal{A}}$ and worlds v. Now assume that i believes in φ at w in $\mathcal{M}$, which means that $V_{\mathfrak{M}_{\mathcal{P},\mathcal{A}},w}(\Box_i\varphi) = 1$. This entails that for all $v \in R_a(w)$ we have $V_{\mathfrak{M}_{\mathcal{P},\mathcal{A}},v}(\varphi) = 1$. But that cannot be if there is at least one world in $R_a(w)$ (by seriality) and if φ is a contradiction. □

Moreover, we might want to put additional constraints on higher-order beliefs of an agent. A higher-order belief of agent i is a belief of i about their own beliefs, e.g., that Alex believes that Alex considers it possible that φ ($\Box_a\Diamond_a\varphi$). A common requirement for rational higher-order beliefs is *positive introspection*: if Alex believes that φ, then Alex believes that they believe φ. Positive introspection is guaranteed by transitive accessibility relations.

Positive Introspection: for all i: R_i is transitive

Claim 2. If an agent i's accessibility relation is transitive in model $\mathcal{M}$, then $V_{\mathcal{M},w}(\Box_i\varphi \to \Box_i\Box_i\varphi) = 1$ for all w in that model.

Proof. Let $\mathcal{M}$ have a transitive accessibility relation R_i. Suppose towards contradiction that $V_{\mathcal{M},w}(\Box_i\varphi \to \Box_i\Box_i\varphi) = 0$ at some world w. That can only be the case if $V_{\mathcal{M},w}(\Box_i\varphi) = 1$ and $V_{\mathcal{M},w}(\Box_i\Box_i\varphi) = 0$. From $V_{\mathcal{M},w}(\Box_i\Box_i\varphi) = 0$ we infer that there is a $w' \in R_i(w)$ such that there is some $w'' \in R_i(w')$ such that $V_{\mathcal{M},w''}(\varphi) = 0$. Since, by R_i is assumed to be transitive, we derive that $w'' \in R_i(w)$. But that contradicts $V_{\mathcal{M},w}(\Box_i\varphi) = 1$, which we have derived previously. □

A final requirement on rational higher-order beliefs, which is often made but more controversial, is that rational agents should also have *negative introspection*: if Alex does not believe that φ, then Alex believes that they don't believe φ. Negative introspection is guaranteed by Euclidean accessibility relations. A relation $R \subseteq W \times W$ is Euclidean iff for all $w, u, v \in W$: if wRu and wRv, then uRv.

Negative Introspection: for all i: R_i is Euclidean

Claim 3. If an agent i's accessibility relation is Euclidean in model $\mathcal{M}$, then $V_{\mathcal{M},w}(\neg\Box_i\varphi \to \Box_i\neg\Box_i\varphi) = 1$ for all w in that model.

Proof. Let $\mathcal{M}$ have a Euclidean accessibility relation R_i. Suppose towards contradiction that $V_{\mathcal{M},w}(\neg\Box_i\varphi \to \Box_i\neg\Box_i\varphi) = 0$ at some world w. That can only be the case if $V_{\mathcal{M},w}(\neg\Box_i\varphi) = 1$ and $V_{\mathcal{M},w}(\neg\Box_i\Box_i\varphi) = 0$. From the former we learn that there is a world $w^* \in R_i(w)$ such that $V_{\mathcal{M},w^*}(\varphi) = 0$. From the latter we learn that for all world $w' \in R_i(w)$ it holds that $V_{\mathcal{M},w'}(\Box_i\varphi) = 1$. But if the relation is Euclidean, then any world $w' \in R_i(w)$ must also "see" w^*, so that is a contradiction. □

Exercise 1. Translate the following sentences into epistemic modal logic. Use as your translation key: p "the earth is round", q "the moon is made of cheese", "a" Alex, and b "Bo."

(i) If Alex believes that the moon is made of cheese, then Alex also believes that the earth is not round.

(ii) Bo believes that Alex thinks it's possible that Bo believes that the moon is made of cheese.

(iii) Alex considers it possible that, if the moon is made of cheese, then Bo believes it.

(iv) Whatever Bo believes about whether the earth is round, Alex believes the same.

(v) Alex is uncertain about whether Bo believes that Alex believes that the moon is made of cheese.

Exercise 2. Here is a full mathematical specification of a modal logical language and a modal model $\mathcal{M}$:

$$\mathcal{P} = \{p, q\}, \quad \mathcal{A} = \{a, b\}, \quad W = \{w_1, w_2, w_3, w_4\}$$

$$V(w_1, p) = V(w_1, q) = 1, \quad V(w_2, p) = 1, V(w_2, q) = 0$$

$$V(w_3, p) = 0, V(w_3, q) = 1, \quad V(w_4, p) = 1, V(w_4, q) = 1$$

$$R_a = \{\langle w_1, w_2\rangle, \langle w_2, w_3\rangle, \langle w_4, w_3\rangle\}$$

$$R_b = \{\langle w_1, w_4\rangle, \langle w_2, w_4\rangle, \langle w_3, w_4\rangle, \langle w_4, w_1\rangle, \langle w_4, w_4\rangle\}$$

1. Draw the model as a diagram like in Figure 2.
2. Determine the truth value for the following:

   a $V_{\mathcal{M},w_1}(\Box_a p)$

   b $V_{\mathcal{M},w_1}(\Box_a p \wedge \Box_a \Diamond_a \neg p)$

   c $V_{\mathcal{M},w_3}(\Box_a (p \wedge \neg p))$

   d $V_{\mathcal{M},w_4}(p \rightarrow \Box_b p)$

3. Does this model satisfy the constraints on rational (higher-order) beliefs? What minimal changes to the accessibility relations would make it satisfy these constraints? Which of the previous truth-values in pointed models change when you look at rational beliefs?

3 Excursion: Applications of modal logic in linguistics and beyond

Beyond a formalization of (rational) beliefs and means to fix the truth-conditions for statements about them, epistemic modal logics have many other applications for linguistics and other neighboring fields.

3.1 Strict conditionals

The truth-conditional semantics for the material conditional $p \rightarrow q$ does not square very well with common intuitions about the meaning of natural language "if ..., then ..." sentences. If Jones says "It's false that if p, then q," they are not saying that p is true and q is false, but that is exactly what the truth-table for $p \rightarrow q$ would give us.

An alternative analysis for natural language conditionals, which is maybe better (but still not unproblematic), is that of the *strict conditional*. According to a strict conditional analysis, we would translate "if p, then q" to something like $\Box_i(p \rightarrow q)$, where i is the speaker. So, we would treat a conditional sentence as (possibly: implicitly) modalized; a statement about the ways the world could be that the speaker considers possible. If Jones says "It's false that if p, then q," this would be translated as $\neg\Box_j(p \rightarrow q)$, which is much weaker than before, as it only requires that Jones thinks that it is conceivable that p is true but q is false.

3.2 Common ground

In communication, we must keep track of who knows what. Otherwise, we risk redundancy and boredom or miscommunication and frustration. Leading theories of pragmatic language use therefore assume that interlocutors keep track of a *common ground* which (among other things) contains the shared beliefs of all participants in a conversation. Modal logics of belief help define such inter-personal notions like *mutual belief* or *common belief*.