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. $(\varphi \rightarrow \psi)$ 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):

$\Box_a(p \rightarrow q)$	Alex believes that if p, then q.
$p \rightarrow \Diamond_a\Box_b q$	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 as live possibilities all the words 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.

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

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 the sake of a simple example, let's consider three possible worlds, $W = \{w_r, w_c, w_s\}$. We need to define a valuation function that maps every pair consisting of a possible world and a proposition letter onto a truth value. Here is (a 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

[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).

Let's furthermore assume the following accessibility relation:

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

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.

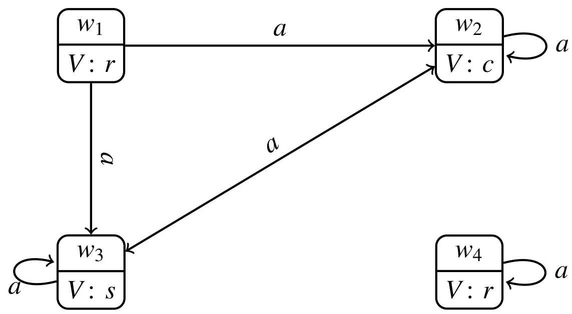

Figure 1: Example of a modal model for Alex's beliefs about tomorrow's weather. In world w_r Alex (falsely) believes that is will not rain.

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 it will not rain. 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:

$$R_a(w_r) = R_a(w_c) = R_a(w_s) = \{w_c, w_s\}$$

So, in these worlds, Alex considers two worlds to be possibly true (namely w_2 and w_3), thus considering cloudy and sunny weather live possibilities but not rain. In w_1 these beliefs are false, in w_2 and w_3 they are not. Finally, in world w_4 Alex's accessible worlds are $R_a(w_4) = \{w_4\}$. So, in w_4, Alex believes truthfully (we might say: "Alex knows") that is will rain.

Notation and terminology. In everything that follows, we implicitly restrict attention to what are called *serial models*, where $R_i(w) \neq \emptyset$ for all w and i. We say that a model is reflexive (r), transitive (t) or Euclidean (e) if all of its relations R_i are reflexive, transitive or Euclidean. For $C \subseteq \{r, t, e\}$ a set of properties, let $\mathfrak{M}^C_{\mathcal{P},\mathcal{A}}$ refer to the collection of all models $\mathcal{M}_{\mathcal{P},\mathcal{A}}$ that have the properties selected in C. Specifically:

$\mathfrak{M}^{\{t,e\}}_{\mathcal{P},\mathcal{A}}$ are *belief models*, and

$\mathfrak{M}^{\{r,t,e\}}_{\mathcal{P},\mathcal{A}}$ are *knowledge models*.

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. Truth of a formula in a pointed model $\mathcal{M}, w$ is defines as follows. For proposition letters $p \in \mathcal{P}$, the valuation function for world w decides on truth or falsity:

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

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

$$\begin{array}{ll} \mathcal{M}, w \models \neg\varphi & \text{iff } \mathcal{M}, w \not\models \varphi \\ \mathcal{M}, w \models \varphi \wedge \psi & \text{iff } \mathcal{M}, w \models \varphi \text{ and } \mathcal{M}, w \models \psi \\ \vdots & \vdots \end{array}$$

What is new is the treatment of modal operators. For these, the accessibility relations are important.[3]

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

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

For example, the model shown in Figure 2 makes the following formulas true:

$\Box_j(\neg p \wedge \neg q)$
"j believes that p and q are false"

$\neg p \wedge \Box_i p$
"i falsely believes that p is true"

$\neg\Box_i q \wedge \neg\Box_i\neg q$
"i is uncertain about q"

$\neg\Box_j p \wedge \Box_i\Box_j p$
"i falsely believes that j believes p"

$\neg\Box_i\Box_j q \wedge \neg\Box_i\neg\Box_j q$
"i is uncertain whether j believes q"

$\Box_i(q \to \Box_j q \wedge \neg q \to \Box_j\neg q)$
"i believes that j knows whether q"

Figure 2 below shows an example of a modal model for two agents ($\mathcal{A} = \{a, b\}$) and just two propositions letters ($\mathcal{P} = \{p, q\}$). This model only has three possible worlds. 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. Concretely, agent a "sees" worlds v and u in w (formally: $R_a(w) = \{u, v\}$), which is shown in the diagram in exactly two arrows with label a "leaving" world w.

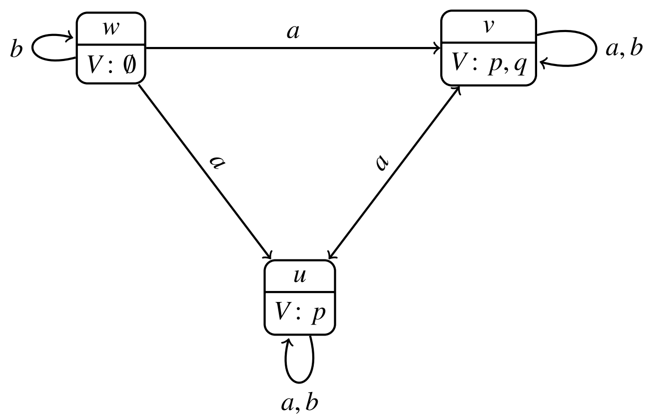

Figure 2: Example of a modal model with three worlds (w, v and u) for two agents (a and b) and two proposition letters (p and q).