AI2Bh Modal Logic

Michael Fourman
Revised by Alan Bundy

May 12, 2004

Modal logics were introduced to formalise *modalities* such as necessity and possibility.

Syntax The, now standard, syntax for these modalities is

$$\Box\varphi \text{ to express, } \varphi \textit{ is necessarily true}$$
$$\Diamond\varphi \text{ to express, } \varphi \textit{ is possibly true}$$

These are normally considered definable in terms of each other:

$$\Diamond\varphi \Leftrightarrow \neg\Box\neg\varphi$$

We will use modal logic to formalise the assertion that an agent, A, *knows that* φ. Knowledge is formally similar to neccessity, so we use the (reasonably standard) syntax,

$$\boxed{\mathsf{K}_A}\,\varphi \text{ to express, } A \textit{ knows that } \varphi$$

Example We can use this syntax to express complex statements about knowledge, and make subtle differences explicit and precise. For example,

$$\exists x.\,\boxed{\mathsf{K}_A}\,\varphi(x) \text{ means, } \textit{for some } x,\ A \textit{ knows that } \varphi(x)$$
$$\boxed{\mathsf{K}_A}\,\exists x.\,\varphi(x) \text{ means, } A \textit{ knows that, for some } x, \varphi(x)$$

Suppose $\varphi(x)$ means, *x is the name of the oldest person in Edinburgh,* and you are A. Which of the two statements is true? (Possibly both are, but that is unlikely.)

Semantics The semantics we have in mind is based on a notion of possible worlds. The idea is that in addition to the actual world, which provides a standard interpretation of predicate logic, we consider other possible worlds. Of all the possibilities that could occur, only some are compatible with A's knowledge – but these are all, from A's perspective, equivalent. So, given a world, W, for each agent, A, there is a set of worlds equivalent, from A's perspective, to W. Formally, we have a set of possible worlds, and for each agent, A, an equivalence relation $\equiv_A$. Informally, we interpret $\boxed{\mathsf{K}_A}\,\varphi$ to mean that φ holds in all the worlds possible for A. Formally,

$$W \models \boxed{\mathsf{K}_A}\,\varphi \textbf{ iff } \text{for all } V \text{ such that } V \equiv_A W, \text{ we have } W \models \varphi$$

Implicitly, this attributes to A omniscience: infallible, and exhaustive powers of reasoning, since

$$\text{if } \boxed{\mathsf{K}_A}\,\varphi \text{ and } \varphi \models \psi \text{ then } \boxed{\mathsf{K}_A}\,\psi$$

As a model of human knowledge, this is clearly flawed; nevertheless, it gives a useful system that approximates some aspects of knowledge as we know it.

Inference We use the possible world semantics to justify the validity of some reasoning principles for modal logic.

Suppose that in all an agent, A's possible worlds, both $\varphi \rightarrow \psi$ and φ hold—*i.e.* suppose $\boxed{\mathsf{K}_A}(\varphi \rightarrow \psi)$, and $\boxed{\mathsf{K}_A}\varphi$ hold, actually. Then in each of these worlds ψ holds—which means that $\boxed{\mathsf{K}_A}\psi$. So $\boxed{\mathsf{K}_A}(\varphi \rightarrow \psi) \wedge \boxed{\mathsf{K}_A}\varphi \rightarrow \boxed{\mathsf{K}_A}\psi$ holds. Equivalently:

$$W \models \boxed{\mathsf{K}_A}(\varphi \rightarrow \psi) \rightarrow \Big(\boxed{\mathsf{K}_A}\varphi \rightarrow \boxed{\mathsf{K}_A}\psi\Big) \qquad \text{(K)}$$

The actual world is viewed as possible by all agents, and whatever any agent knows is, in fact, *i.e. in the actual world*, true. Formally, $\equiv_A$ is reflexive: $W \equiv_A W$.

So, if $W \models \boxed{\mathsf{K}_A}\varphi$ **then we have** $W \models \varphi$**.**

$$W \models \boxed{\mathsf{K}_A}\varphi \rightarrow \varphi \qquad \text{(T)}$$

Furthermore, from the perspective of any agent, all its possible worlds are equivalent, so whatever it knows in the actual world, it knows in all its possible worlds — which means that, actually, it knows that it knows. Formally, since $\equiv_A$ is an equivalence relation, it is transitive: if $V \equiv_A W$ then, for each $U \equiv_A V$ we have $U \equiv_A W$.

So, if $W \models \boxed{\mathsf{K}_A}\varphi$ **and** $V \equiv_A W$ **then**
for each $U \equiv_A V$ **we have** $U \models \varphi$
which means that $V \models \boxed{\mathsf{K}_A}\varphi$
and so, $W \models \boxed{\mathsf{K}_A}\boxed{\mathsf{K}_A}\varphi$**.**

$$W \models \boxed{\mathsf{K}_A}\varphi \rightarrow \boxed{\mathsf{K}_A}\boxed{\mathsf{K}_A}\varphi \qquad \text{(4)}$$

Finally, $\equiv_A$ is symetric: if $V \equiv_A W$ then $W \equiv_A V$.

If, $W \models \neg\boxed{\mathsf{K}_A}\varphi$ **then, for some** $V \equiv_A W$**, we have** $V \not\models \varphi$**.**
Now, for any $U \equiv_A W$**, we have** $V \equiv_A U$**, so** $U \not\models \boxed{\mathsf{K}_A}\varphi$**; i.e.** $U \models \neg\boxed{\mathsf{K}_A}\varphi$**.**
Which means that $W \models \boxed{\mathsf{K}_A}\neg\boxed{\mathsf{K}_A}\varphi$**.**

$$W \models \neg\boxed{\mathsf{K}_A}\varphi \rightarrow \boxed{\mathsf{K}_A}\neg\boxed{\mathsf{K}_A}\varphi \qquad \text{(5)}$$

From this discussion, it is clear that if our semantics were defined in terms of an arbitrary relation, in place of the equivalence relation we have used, then different principles could be justified depending on the properties of this relation: **K** would always be valid; **T** if the relation were reflexive; **4** if transitive; **5** if transitive and symmetric. These four, **{K, T, 4, 5}**, corresponding to an equivalence relation, characterise a so-called **S5** modality. The logic for a reflexive relation, satisfying **{K, T}**, is called **KT**. For a reflexive, transitive relation, we have the logic **S4**, satisfying **{K, T, 4}**[1].

[1] **K, KT, S4, S5** are the traditional names for modal logics with a single neccessity operator satisfying these rules; the logics discussed here are multimodal logics.

Common Knowledge Sometimes it is not enough for A to know that B knows, and so on; sometimes we need to know that everybody knows and that everybody knows that everybody knows. We introduce special modalities

$$\boxed{\mathsf{E}}\,\varphi \text{ \textbf{everybody knows} that } \varphi$$
$$\boxed{\mathsf{C}}\,\varphi \text{ it is \textbf{common knowledge} that } \varphi$$

where

$$\boxed{\mathsf{E}}\,\varphi \Leftrightarrow \boxed{\mathsf{K}_A}\,\varphi \wedge \boxed{\mathsf{K}_B}\,\varphi \wedge \ldots$$
$$\boxed{\mathsf{C}}\,\varphi \Leftrightarrow \boxed{\mathsf{E}}\,(\varphi \wedge \boxed{\mathsf{C}}\,\varphi)$$

Here, *everybody knows* is defined by a conjunction, over all agents. *Common knowledge* is defined recursively.

Semantically, the modality $\boxed{\mathsf{E}}$, *everybody knows*, corresponds to the relation which is the union of all the $\equiv_A$. The union of equivalence relations is not, in general, an equivalence relation; it is not transitive. The transitive closure of this relation is an equivalence relation, it provides the model for the *common knowledge* modality, $\boxed{\mathsf{C}}$.

We will characterise speech acts as actions that agents can plan to take, using modal logic to express their preconditions and effects.