IX. Epistemic Logic

AS.150.498: Modal Logic and Its Applications
Johns Hopkins University, Spring 2017

Another important application is *epistemic logic* which is concerned with the individual and collective knowledge states of groups of agents, and how these states change when new information comes to light.

1 Syntax, Semantics, and Proof Systems

We will start off with the following polymodal language:

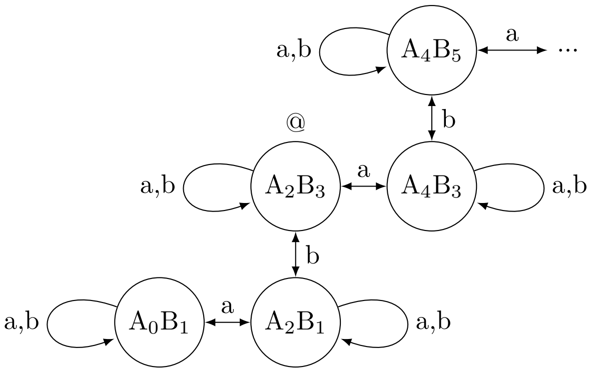
Definition 9.1. The **epistemic language** $\mathcal{L}_e$ extends the basic sentential language with knowledge operators for each agent in set Agt=$\{a, b, c, ...\}$:

$$p \mid \bot \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \mathrm{K}_a\varphi$$

Read $\mathrm{K}_a\varphi$ as 'Agent a knows that φ.' Defining the dual $\mathrm{K}^*_a = \neg\mathrm{K}_a\neg$, read $\mathrm{K}^*_a\varphi$ as 'It is compatible with what Agent a knows that φ.'

A model $\mathcal{M} = \langle \mathcal{W}, \{\mathcal{R}_a\}_{a\in\mathrm{Agt}}, \mathcal{V}\rangle$ for $\mathcal{L}_e$ is a standard Kripke model with an epistemic accessibility relation for each agent in Agt, where $w\mathcal{R}_a v$ just in case v is epistemically possible for Agent a in w—that is, Agent a's knowledge in w leaves open v. (@ will designate the actual world.)

For example, suppose that Agent a has a 2 written on her forehead and Agent b has a 3 written on his forehead. Each agent can see the other's forehead but they do not know the number on their own forehead. They are told by a reliable source that the numbers on their foreheads are n and $n + 1$ for some $n \in \mathbb{N}$.



where A_n designates that Agent a has n on her forehead, and B_n designates that Agent b has n on his forehead. Note that an a-arrow or b-arrow reflects Agent a's or Agent b's *ignorance* respectively.

In this model $\mathcal{M}$, $[\![\mathrm{K}_a B_3 \wedge \mathrm{K}_b A_2]\!]^@_\mathcal{M} = T$, $[\![\mathrm{K}_a\neg B_5 \wedge \mathrm{K}_b\neg B_5]\!]^@_\mathcal{M} = T$, but $[\![\mathrm{K}_a\mathrm{K}_b\neg B_5]\!]^@_\mathcal{M} = F$.

Since knowledge is factive, the **T** axiom $\mathrm{K}_a\varphi \supset \varphi$ is valid (hence the reflexive loops for each agent at each world in the above model—these are typically left implicit).

The validity of other axioms is more controversial. The **4** axiom is the famous KK-principle $\mathrm{K}_a\varphi \supset \mathrm{K}_a\mathrm{K}_a\varphi$ (also known as *positive introspection*). We will soon discuss an argument of Williamson against this principle.

While philosophers reject **5** (also known as *negative introspection*) and **B**, economists and computer scientists typically assume that the logic of knowledge is **S5**.

2 Collective Knowledge

So far, we have considered only the knowledge of individual agents. But we can also define some interesting notions of collective knowledge. For instance, we might introduce the following *everyone in X knows* operator E_X (where $X \subseteq \mathrm{Agt}$):

$$[\![\mathrm{E}_X\varphi]\!]^w_\mathcal{M} = T \quad \text{iff} \quad \forall a \in X([\![\mathrm{K}_a\varphi]\!]^w_\mathcal{M} = T)$$

If $|X|$ is finite, then E_X is clearly definable in $\mathcal{L}_e$: $\mathrm{E}_X\varphi \equiv \bigwedge_{a\in X}\mathrm{K}_a\varphi$.

We might also introduce this *someone in X knows* operator S_X:

$$[\![\mathrm{S}_X\varphi]\!]^w_\mathcal{M} = T \quad \text{iff} \quad \exists a \in X([\![\mathrm{K}_a\varphi]\!]^w_\mathcal{M} = T)$$

If $|X|$ is finite, then $\mathrm{S}_X\varphi \equiv \bigvee_{a\in X}\mathrm{K}_a\varphi$.

More interestingly, we can introduce the notions of **common knowledge** and **distributed knowledge**. Something is common knowledge among a group of agents X iff everyone in X knows it and everyone in X knows that everyone in X knows it, and so forth:[1]

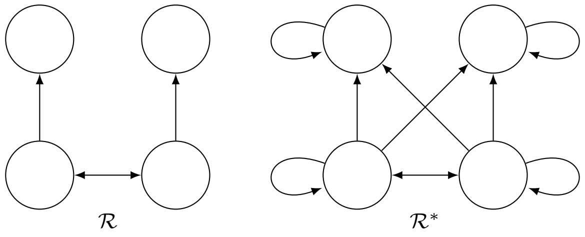
$$[\![\mathrm{C}_X\varphi]\!]^w_\mathcal{M} = T \quad \text{iff} \quad [\![\mathrm{E}_X\varphi]\!]^w_\mathcal{M} = [\![\mathrm{E}_X\mathrm{E}_X\varphi]\!]^w_\mathcal{M} = ... = T$$

Our Kripke models afford a more elegant truth clause:

[1] Adding C_X to $\mathcal{L}_e$ significantly increases expressive power. We can then define finite pointed models up to bisimulation—that is, for each finite $\mathcal{M}, w$ there is some sentence φ in the expanded language such that $[\![\varphi]\!]^v_\mathcal{N} = T$ iff $\mathcal{M}, w \underline{\leftrightarrow} \mathcal{N}, v$.

$$[\![\mathrm{C}_X\varphi]\!]^w_{\mathcal{M}} = T \quad \text{iff} \quad \text{For all } v \in \mathcal{W}, \text{ if } v \text{ is reachable from } w \text{ in a finite number of steps along any } \mathcal{R}_a \text{ where } a \in X, \text{ then } [\![\varphi]\!]^v_{\mathcal{M}} = T$$

Given a relation $\mathcal{R}$, let $\mathcal{R}^*$ be the reflexive transitive closure of $\mathcal{R}$—that is, $\mathcal{R}^*$ is the relation obtained from $\mathcal{R}$ by adding reflexive loops and whatever is required for transitivity.



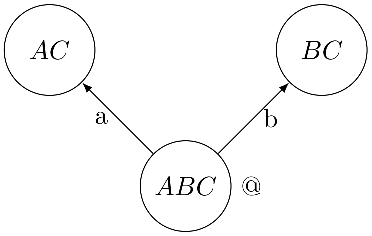
Then the truth clause for C_X can be restated thus:

$$[\![\mathrm{C}_X\varphi]\!]^w_{\mathcal{M}} = T \quad \text{iff} \quad \forall v \in \{v : w(\textstyle\bigcup_{a\in X} \mathcal{R}_a)^* v\}([\![\varphi]\!]^v_{\mathcal{M}} = T)$$

By contrast, something is distributed knowledge among a group of agents X iff, roughly, the agents would know it were they to share all of their individual knowledge:

$$[\![\mathrm{D}_X\varphi]\!]^w_{\mathcal{M}} = T \quad \text{iff} \quad \forall v \in \{v : w \textstyle\bigcap_{a\in X} \mathcal{R}_a v\}([\![\varphi]\!]^v_{\mathcal{M}} = T)$$

Informally, v is an epistemic possibility post-sharing in w just in case v is epistemically possible for each member of X in w. If any agent's individual knowledge rules out v, then the agents' distributed knowledge will also rule out v.



In this model $\mathcal{M}$, $[\![\mathrm{E}_X(A \vee B)]\!]^@_{\mathcal{M}} = T$, $[\![\mathrm{S}_X A]\!]^@_{\mathcal{M}} = T$, $[\![\mathrm{C}_X C]\!]^@_{\mathcal{M}} = T$, and $[\![\mathrm{D}_X(A \wedge B)]\!]^@_{\mathcal{M}} = T$.

3 Against KK

Before turning to dynamic epistemic logic, let us first briefly consider an argument of Williamson [2000] against the KK-principle.

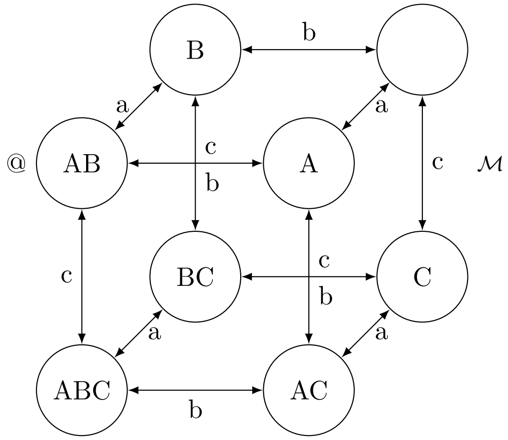
Suppose that you are looking at a tree in the distance. If H_n designates that the tree is n inches tall, then H_{1000}. Given that your eyesight is imperfect, $\mathrm{K}\neg H_n \supset \neg H_{n+1}$, and you can come to know this by reflecting on your visual limitations. But Williamson argues that KK then leads to trouble:

1. $\mathrm{K}\neg H_{500}$ PL
2. $\mathrm{K}\neg H_{500} \supset \neg H_{501}$ Assumption
3. $\mathrm{K}(\mathrm{K}\neg H_{500} \supset \neg H_{501})$ Assumption
4. $\mathrm{KK}\neg H_{500}$ **4** Axiom 1
5. $\mathrm{KK}\neg H_{500} \supset \mathrm{K}\neg H_{501}$ **K** Axiom 3
6. $\mathrm{K}\neg H_{501}$ MP 5,4

Repeating this reasoning, you can conclude $\mathrm{K}\neg H_{1000}$ and so $\neg H_{1000}$ by the **T** Axiom. This contradicts H_{1000}.

4 Going Dynamic

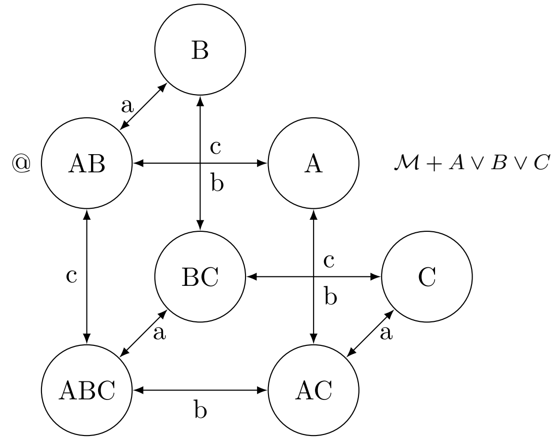
Time to get dynamic. To introduce some of the key ideas of dynamic epistemic logic, let us work through the Muddy Children Puzzle. Three children a, b, and c have been playing outside in the mud. Let A, B, and C designate that a, b, and c have mud on their forehead respectively. In fact, A and B but $\neg C$. Here is the initial epistemic model:



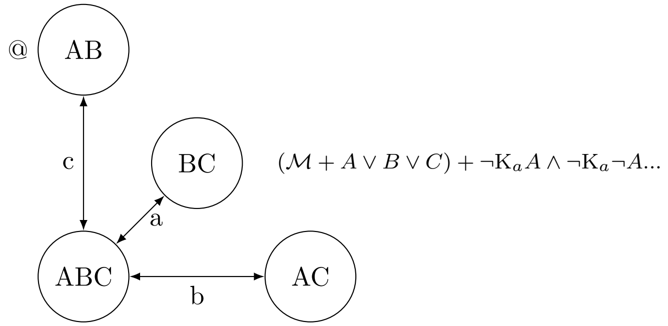
Returning home, their mother says that at least one of the children has mud on their forehead. How does the model change?

Definition 9.2. Given $\mathcal{M} = \langle \mathcal{W}, \{\mathcal{R}_a\}_{a\in \text{Agt}}, \mathcal{V}\rangle$, the **model updated with** φ is $\mathcal{M} + \varphi = \langle \mathcal{W} + \varphi, \{\mathcal{R}_a + \varphi\}_{a\in \text{Agt}}, \mathcal{V} + \varphi\rangle$ where:

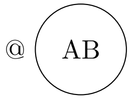
$\mathcal{W} + \varphi = \{w \in \mathcal{W} : [\![\varphi]\!]^w_{\mathcal{M}} = T\}$
$\mathcal{R}_a + \varphi$ is the restriction of $\mathcal{R}_a$ to $\mathcal{W} + \varphi$
$\mathcal{V} + \varphi$ is the restriction of $\mathcal{V}$ to $\mathcal{W} + \varphi$



She then asks each of the children to step forward if they know whether they have mud on their forehead. Since $[\![\neg \text{K}_a A \wedge \neg \text{K}_a \neg A]\!]^{@}_{\mathcal{M}+A\vee B\vee C} = T$, $[\![\neg \text{K}_b B \wedge \neg \text{K}_b \neg B]\!]^{@}_{\mathcal{M}+A\vee B\vee C} = T$, and $[\![\neg \text{K}_c C \wedge \neg \text{K}_c \neg C]\!]^{@}_{\mathcal{M}+A\vee B\vee C} = T$, none of the children step forward. This provides new information.


@
AB
c
BC
$(\mathcal{M} + A \vee B \vee C) + \neg \text{K}_a A \wedge \neg \text{K}_a \neg A ...$
a
ABC
AC
b

The mother again asks each of the children to step forward if they know whether they are dirty. Since $[\![\text{K}_a A]\!]^{@}_{\mathcal{M}+...} = [\![\text{K}_b B]\!]^{@}_{\mathcal{M}+...} = T$, a and b step forward. But since $[\![\neg \text{K}_c C \wedge \neg \text{K}_c \neg C]\!]^{@}_{\mathcal{M}+...} = T$, c does not. Again, this provides new information. In fact, only @ remains open after this last update. So next time the mother asks her question, c steps forward as well.



To talk about what holds after an information update, we can add dynamic operators to the language $\mathcal{L}_e$:

Definition 9.3. The **language of Public Announcement Logic** $\mathcal{L}_{PAL}$ is given by:

$$p \mid \bot \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \text{K}_a\varphi \mid [!\varphi]\varphi \mid \langle !\varphi\rangle\varphi$$

Read $[!\varphi]\psi$ as 'After every true announcement of φ, ψ.' Read $\langle !\varphi\rangle\psi$ as 'After some true announcement of φ, ψ.'

The truth clauses for these dynamic operators are as follows:

$$\begin{array}{lll} [\![[!\varphi]\psi]\!]^w_{\mathcal{M}} = T & \text{iff} & [\![\varphi]\!]^w_{\mathcal{M}} = F \text{ or } [\![\psi]\!]^w_{\mathcal{M}+\varphi} = T \\ [\![\langle !\varphi\rangle\psi]\!]^w_{\mathcal{M}} = T & \text{iff} & [\![\varphi]\!]^w_{\mathcal{M}} = T \text{ and } [\![\psi]\!]^w_{\mathcal{M}+\varphi} = T \end{array}$$

The two operators are duals: $[!\varphi]\psi \equiv \neg\langle !\varphi\rangle\neg\psi$ and $\langle !\varphi\rangle\psi \equiv \neg[!\varphi]\neg\psi$.

The operator $\langle !\varphi\rangle$ is more intuitive. To determine whether $\langle !\varphi\rangle\psi$ is true at w, first check whether φ is true at w. If not, $\langle !\varphi\rangle\psi$ is false. If so, then next check whether ψ is true at w in the model updated with φ. If so, $\langle !\varphi\rangle\psi$ is true. If not, $\langle !\varphi\rangle\psi$ is false.

Using these operators, we can talk about the kind of things that happen in the Muddy Children Puzzle:

$[\![\langle !A \vee B \vee C\rangle(\neg \text{K}_a A \wedge \neg \text{K}_a \neg A)]\!]^{@}_{\mathcal{M}} = T$
$[\![\langle !A \vee B \vee C\rangle(\neg \text{K}_b B \wedge \neg \text{K}_b \neg B)]\!]^{@}_{\mathcal{M}} = T$
$[\![\langle !A \vee B \vee C\rangle(\neg \text{K}_c C \wedge \neg \text{K}_c \neg C)]\!]^{@}_{\mathcal{M}} = T$

Letting $\varphi^* = \neg \text{K}_a A \wedge \neg \text{K}_a \neg A \wedge \neg \text{K}_b B \wedge \neg \text{K}_b \neg B \wedge \neg \text{K}_c C \wedge \neg \text{K}_c \neg C$,

$[\![\langle !A \vee B \vee C\rangle\langle !\varphi^*\rangle \text{K}_a A]\!]^{@}_{\mathcal{M}} = T$
$[\![\langle !A \vee B \vee C\rangle\langle !\varphi^*\rangle \text{K}_b B]\!]^{@}_{\mathcal{M}} = T$
$[\![\langle !A \vee B \vee C\rangle\langle !\varphi^*\rangle(\neg \text{K}_c C \wedge \neg \text{K}_c \neg C)]\!]^{@}_{\mathcal{M}} = T$

$[\![\langle !A \vee B \vee C\rangle\langle !\varphi^*\rangle\langle !\text{K}_a A \wedge \text{K}_b B\rangle \text{K}_c \neg C]\!]^{@}_{\mathcal{M}} = T$

Definition 9.4. The update of $\mathcal{M}, w$ with φ is **successful** if and only if $[\![\langle !\varphi\rangle\varphi]\!]^w_{\mathcal{M}} = T$.

Definition 9.5. The update of $\mathcal{M}, w$ with φ is **unsuccessful** if and only if $[\![\langle !\varphi\rangle\neg\varphi]\!]^{w}_{\mathcal{M}} = T$.

That is, an update with φ is unsuccessful if and only if φ becomes false after it is announced. For instance, the update of $\mathcal{M} + A \vee B \vee C, @$ with φ^* is unsuccessful since $[\![\varphi^*]\!]^{@}_{(\mathcal{M}+A\vee B\vee C)+\varphi^*} = F$.

Definition 9.6. The sentence φ is **successful** if and only if $[!\varphi]\varphi$ is valid.

Definition 9.7. The sentence φ is **unsuccessful** if and only if $[!\varphi]\varphi$ is invalid.

Definition 9.8. The sentence φ is **self-refuting** if and only if $[!\varphi]\neg\varphi$ is valid.

An example of a self-refuting sentence is $A \wedge \neg \mathrm{K}_a A$.

Note that *uniform substitution* fails in PAL. While $[!A]A$ is valid, $[!\varphi]\varphi$ is not.