9 Epistemic Logic

For modelling knowledge in a multi-agent system, we will use multimodal logic, i.e., modal logic with several modalities. This will allow us to talk about the relationship between possible worlds from different aspects (viz. according to the knowledge of agents).

9.1 Syntax and Semantics

The language of the multimodal logic $\mathbf{S5_n}$ (n a natural number) is defined as in the case of basic modal logic, but we have n different boxes, denoted as K_i ($i \leq n$). That is,

- for each $i \leq n$, $K_i\varphi$ is a formula for each formula φ.

The intuition is that $K_i\varphi$ holds at a possible world s if "in s, agent i knows φ" — in other words, φ is true in all worlds that agent i considers possible. That is why in the semantics, there are n accessibility relations $\mathcal{K}_i$ (for each agent $i \leq n$): $s\mathcal{K}_i t$ holds if agent i considers t a possible alternative of s, i.e., if s and t are indistinguishable according to agent i's knowledge. Hence we require that $\mathcal{K}_i$ is an equivalence relation (reflexive, symmetric and transitive).

More formally, a frame is $\mathcal{F} = (S, \mathcal{K}_i : i \leq n)$ where S is a non-empty set and each $\mathcal{K}_i \subseteq S \times S$ is an equivalence relation. A model is a frame together with a valuation: $\mathcal{M} = (\mathcal{F}, V)$. Truth in a model is defined in the usual way:

- $(\mathcal{M}, s) \models K_i\varphi$ iff $(\mathcal{M}, t) \models \varphi$ for all t such that $s\mathcal{K}_i t$.

9.2 Adding Common and Distributed Knowledge

Next we expand the language to incorporate common and distributed knowledge.

Given a collection $G \subseteq \{1, \ldots, n\}$,

- $E_G\varphi$, $C_G\varphi$ and $D_G\varphi$ are formulas whenever φ is a formula.

$E_G\varphi$ expresses that "every agent $i \in G$ knows φ", i.e.,

- $(\mathcal{M}, s) \models E_G\varphi$ iff $(\mathcal{M}, s) \models K_i\varphi$ for all $i \in G$.

Let $E_G^0\varphi$ stand for φ and $E_G^{k+1}\varphi$ for $E_G^k E_G\varphi$. Intuitively, "φ is common knowledge in group G" if everyone in G knows that everyone in G knows that φ, etc. That is,

- $(\mathcal{M}, s) \models C_G\varphi$ iff $(\mathcal{M}, s) \models E_G^k\varphi$ for all $k = 1, 2, \ldots$.

Finally, we can combine the knowledge of the agents in group G by eliminating those worlds that some agent in G considers impossible (and thus restricting the number of alternatives):

- $(\mathcal{M}, s) \models D_G\varphi$ iff $(\mathcal{M}, t) \models \varphi$ for all t such that $(s, t) \in \bigcap_{i \in G} \mathcal{K}_i$.

Let us fix a model $(S, \mathcal{K}_i, V : i \leq n)$ and let $s, t \in S$. Given a set $G \subseteq \{1, \ldots, n\}$, we say that *$t$ is G-reachable from s* if there is a path $s = r_0, r_1, \ldots, r_k = t$ such that for each $0 \leq j < k$, there exists $i \in G$ with $(r_j, r_{j+1}) \in \mathcal{K}_i$. Then the following hold.

Lemma 9.1 *1. $(\mathcal{M}, s) \models E_G^k\varphi$ iff $(\mathcal{M}, t) \models \varphi$ for all t that are G-reachable from s in k steps.*

2. $(\mathcal{M}, s) \models C_G\varphi$ iff $(\mathcal{M}, t) \models \varphi$ for all t that are G-reachable from s.

Proof: We prove 1 by induction on k. The base case $k = 0$ is obvious.

In the inductive step, first we assume that $(\mathcal{M}, s) \models E_G^{k+1}\varphi$. Let t be such that it is G-reachable from s in $(k+1)$ steps; we have to show that $(\mathcal{M}, t) \models \varphi$. By the induction hypothesis, $(\mathcal{M}, u) \models E_G\varphi$ for every u that is G-reachable from s in k steps. In particular, $(\mathcal{M}, v) \models E_G\varphi$ for v that is the kth node on the path leading to t from s. It follows that $(\mathcal{M}, t) \models \varphi$.

Next assume that, for every t that it is G-reachable from s in $(k+1)$ steps, we have $(\mathcal{M}, t) \models \varphi$. Then $(\mathcal{M}, u) \models E_G\varphi$ for every u that it is G-reachable from s in k steps; otherwise there were t G-reachable from s in $(k+1)$ steps with $(\mathcal{M}, t) \models \neg\varphi$. By applying the induction hypothesis, we have $(\mathcal{M}, s) \models E_G^k(E_G\varphi)$. That is, $(\mathcal{M}, s) \models E_G^{k+1}\varphi$ as desired.

Since $C_G\varphi$ means that $E_G^k\varphi$ for all k, 2 follows. ∎

9.3 Properties of Knowledge

Here we list some important properties of knowledge. The validity of these properties follow from the way we defined semantics.

- Distribution
$$\models (K_i\varphi \wedge K_i(\varphi \to \psi)) \to K_i\psi$$
- Knowledge generalization
$$\mathcal{M} \models \varphi \text{ implies } \mathcal{M} \models K_i\varphi$$
- Knowledge axiom (cf. belief)
$$\models K_i\varphi \to \varphi$$
- Positive introspection
$$\models K_i\varphi \to K_iK_i\varphi$$
- Negative introspection
$$\models \neg K_i\varphi \to K_i\neg K_i\varphi$$
- A syntactic definition of “everybody knows”
$$\mathcal{M} \models E_G\varphi \leftrightarrow \bigwedge_{i\in G} K_i\varphi$$
- Fixed-point axiom
$$\mathcal{M} \models C_G\varphi \leftrightarrow E_G(\varphi \wedge C_G\varphi)$$
- Induction rule
$$\mathcal{M} \models \varphi \to E_G(\varphi \wedge \psi) \text{ implies } \mathcal{M} \models \varphi \to C_G\psi$$
- For $G' \subseteq G$,
$$\models C_G\varphi \to C_{G'}\varphi \qquad \models D_{G'}\varphi \to D_G\varphi$$

9.4 Examples

Muddy children: Imagine n children playing together. Some of the children get mud on their forehead during the game. They can see other children’s forehead, but not their own forehead. After the game, the father announces to them that “At least one of you have muddy forehead”. Then the father repeatedly asks: “Raise your hand if you know whether your forehead is muddy”.

Exercise 9.2 Using epistemic logic figure out what happens in the muddy children puzzle. Assume that all children are healthy, truthful and intelligent.

Three wise men: There are three wise men. It is common knowledge that there are three red hats and two white hats. The king puts a hat on the head of each of the three wise men, and asks them (sequentially) if they know the colour of the hat on their head. The first wise man says that he does not know; the second wise man says that he does not know; then the third wise man says that he does know.

Exercise 9.3 Using epistemic logic figure out the colour of the third wise man's hat.

10 Temporal Epistemic Logic

In many applications of temporal logic, it is enough to consider special models. In particular, in this section we fix the underlining serial Kripke frame to be $(\mathbb{N}, \mathrm{suc})$, where suc is the successor function $\mathrm{suc}(i) = i + 1$.

We can define a combination of epistemic and temporal logics as follows. The syntax is defined by considering propositional, epistemic and temporal connectives (e.g., $K\mathsf{F}\varphi$ is a formula as well, expressing that "we know (now) that φ will be true sometimes in the future"). The semantics is defined by taking "products" of epistemic (horizontal direction) and temporal (vertical direction) frames, see Figure 11.

Consider the following puzzle: On Friday the teacher says to the students: "There will be an exam next week, but you will not know in advance on which day." One of the students argues as follows: "The exam cannot be on next Friday, since we would know it by Thursday night. But the exam cannot be on Thursday either, since we would know this by Wednesday night (we eliminated Friday above, so the only possibility would be Thursday). Similarly, we can eliminate the other days of the week... Therefore, we cannot have an exam next week."

Let us try to use a combination of epistemic and temporal logics to solve this puzzle. First let us make the teacher's announcement more precise (we could deal with other interpretations in a similar way): "There will be a unique (single) exam next week, but you will not know the date of the exam on the previous date." Let p be an atomic proposition expressing "there is an exam today". Then $\mathsf{X}p$ expresses that "there will be an exam tomorrow". The formula $\neg K\mathsf{X}p$ expresses that "[you] do not know that the exam will be tomorrow". Hence $\mathsf{X}p \wedge \neg K\mathsf{X}p$ expresses that "there will be an exam tomorrow, but you do not know it". Thus $\mathsf{XX}(\mathsf{X}p \wedge \neg K\mathsf{X}p)$ expresses that "there will be an exam on the third day, but you will not know this the day after tomorrow" (i.e., the exam will be on Monday, but you will not know this on Sunday, provided that today is Friday). Similarly $\mathsf{XXX}(\mathsf{X}p \wedge \neg K\mathsf{X}p)$ expresses that "there will be an exam on the fourth day, but you will not know this on the third day", etc. Let $\oplus$ denote exclusive or and X^n denote the sequence of n copies of X. Thus the teacher's statement can be expressed as

$$\bigoplus\{\mathsf{X}^n(\mathsf{X}p \wedge \neg K\mathsf{X}p) : 2 \leq n \leq 6\}.$$

Can we find a model that satisfies this formula?

Consider the model on Figure 11. Here we have five temporal models according to the five possible dates for the exam, plus one in which no exam happens at all (the student cannot be sure whether the teacher said the truth); the date of the exam is indicated by bold font. We "join" them by according to the student's knowledge: the horizontal lines between dates indicate that the student cannot tell these "worlds" apart — these are indistinguishable according to the student's knowledge. We see that the student would not know the date of the exam on the previous day. For instance, if the exam is on Wednesday, then four Tuesdays (when the exam did not happen till Tuesday) are indistinguishable to the student, i.e., $\neg K\mathsf{X}p$ holds at these worlds.

But what about the student's argument? The above argument applies even when the exam is on Friday: since Th_0 and Th_1 are indistinguishable, the student cannot conclude that the exam must be on Friday if it did not happen by Thursday. This shows that the teacher's announcement is satisfiable, viz. on whichever date the exam is, the students would not know the date in advance.

$$
\begin{tikzcd}
{} \arrow[d, dash] & {} \arrow[d, dash] & {} \arrow[d, dash] & {} \arrow[d, dash] & {} \arrow[d, dash] & {} \arrow[d, dash] \\
\mathrm{Fr}_0 \arrow[d, dash] & \mathbf{Fr_1} \arrow[d, dash] & \mathrm{Fr}_2 \arrow[d, dash] & \mathrm{Fr}_3 \arrow[d, dash] & \mathrm{Fr}_4 \arrow[d, dash] & \mathrm{Fr}_5 \arrow[d, dash] \\
\mathrm{Th}_0 \arrow[r, dash] \arrow[d, dash] & \mathrm{Th}_1 \arrow[d, dash] & \mathbf{Th_2} \arrow[d, dash] & \mathrm{Th}_3 \arrow[d, dash] & \mathrm{Th}_4 \arrow[d, dash] & \mathrm{Th}_5 \arrow[d, dash] \\
\mathrm{We}_0 \arrow[r, dash] \arrow[d, dash] & \mathrm{We}_1 \arrow[r, dash] \arrow[d, dash] & \mathrm{We}_2 \arrow[d, dash] & \mathbf{We_3} \arrow[d, dash] & \mathrm{We}_4 \arrow[d, dash] & \mathrm{We}_5 \arrow[d, dash] \\
\mathrm{Tu}_0 \arrow[r, dash] \arrow[d, dash] & \mathrm{Tu}_1 \arrow[r, dash] \arrow[d, dash] & \mathrm{Tu}_2 \arrow[r, dash] \arrow[d, dash] & \mathrm{Tu}_3 \arrow[d, dash] & \mathbf{Tu_4} \arrow[d, dash] & \mathrm{Tu}_5 \arrow[d, dash] \\
\mathrm{Mo}_0 \arrow[r, dash] \arrow[d, dash] & \mathrm{Mo}_1 \arrow[r, dash] \arrow[d, dash] & \mathrm{Mo}_2 \arrow[r, dash] \arrow[d, dash] & \mathrm{Mo}_3 \arrow[r, dash] \arrow[d, dash] & \mathrm{Mo}_4 \arrow[d, dash] & \mathbf{Mo_5} \arrow[d, dash] \\
\mathrm{Su}_0 \arrow[r, dash] & \mathrm{Su}_1 \arrow[r, dash] & \mathrm{Su}_2 \arrow[r, dash] & \mathrm{Su}_3 \arrow[r, dash] & \mathrm{Su}_4 \arrow[r, dash] & \mathrm{Su}_5
\end{tikzcd}
$$

Figure 11: A temporal epistemic model