Tik-79.192 Special course in theoretical computer science, autumn 1998

Lecture notes
by
Patrik Simons

August 25, 1998

1 The language

We assume the existence of a group of agents $1, \ldots, n$ and a set of primitive propositions Φ. The language is then the set of all formulas, where a formula is a primitive proposition or constructed from other formulas using negation ($\neg$), conjunction ($\wedge$) or a modal operator (K_i for $i = 1, \ldots, n$).

Definition 1.1. Let φ be a formula iff

$$\begin{array}{ll} \varphi = p & \text{for } p \in \Phi, \\ \varphi = \neg\psi & \text{for a formula } \psi, \\ \varphi = (\psi \wedge \chi) & \text{for formulas } \psi \text{ and } \chi, \text{ or} \\ \varphi = K_i\psi & \text{for a formula } \psi \text{ and } i = 1, \ldots, n. \end{array}$$

In addition, we use the abbreviations

$$\begin{array}{rl} (\varphi \vee \psi) & \text{for } \neg(\neg\varphi \wedge \neg\psi), \\ \varphi \Rightarrow \psi & \text{for } \neg\varphi \vee \psi, \\ \varphi \Leftrightarrow \psi & \text{for } (\varphi \Rightarrow \psi) \wedge (\psi \Rightarrow \varphi), \\ \top & \text{for } (p \vee \neg p) \text{ for a fixed } p \in \Phi, \text{ and} \\ \bot & \text{for } \neg\top. \end{array}$$

To make the syntax easier to read, we leave out parentheses when doing so does not cause any ambiguities.

Having defined the syntax we can now define the meaning or semantics of a formula. For this we use a formal model: the Kripke structure.

Definition 1.2. A Kripke structure

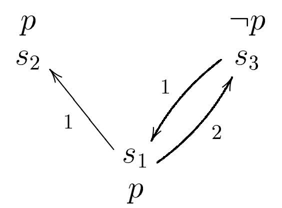
$$M = (S, \pi, \mathcal{K}_1, \ldots, \mathcal{K}_n)$$

for n agents over Φ is made up of a set S of states or possible worlds, an interpretation

$$\pi : S \rightarrow (\Phi \rightarrow \{true, false\})$$

assigning to every state a truth assignment over the primitive propositions, and one binary relation $\mathcal{K}_i$ over S for every agent $i = 1, \ldots, n$.

Example 1.3. Since the binary relations can be thought of as a labeled graph over the possible worlds, we can make a visualization of the structure. For instance,



is the Kripke structure

$$M = (\{s_1, s_2, s_3\}, \pi, \{\langle s_1, s_2\rangle, \langle s_3, s1\rangle\}, \{\langle s_1, s_3\rangle\})$$

when

$$\begin{aligned} \pi(s_1)(p) &= true \\ \pi(s_2)(p) &= true \\ \pi(s_3)(p) &= false \end{aligned}$$

We write

$$(M, s) \models \varphi$$

to denote that the formula φ is true at state s in the Kripke structure M, and we write $(M, s) \not\models \varphi$ to denote that it is false.

Definition 1.4. For the primitive proposition $p \in \Phi$, the formulas φ and ψ, the Kripke structure $M = (S, \pi, \mathcal{K}_1, \ldots, \mathcal{K}_n)$ and the state $s \in S$

$$\begin{array}{lll} (M,s) \models p & \text{iff} & \pi(s)(p) = true \\ (M,s) \models \varphi \wedge \psi & \text{iff} & (M,s) \models \varphi \text{ and } (M,s) \models \psi \\ (M,s) \models \neg\varphi & \text{iff} & (M,s) \not\models \varphi \\ (M,s) \models K_i\varphi & \text{iff} & (M,t) \models \varphi \text{ for all } t \text{ such that } \langle s,t\rangle \in \mathcal{K}_i \end{array}$$

The modal operator K_i has the intuition "agent i knows", where the agent i knows φ in state s if φ is true in all states reachable in one step from s. This is why we also call the states possible worlds. An agent only knows things that are true in all worlds it considers possible.

Example 1.5. Agent 1 does not know that agent 2 knows that agent 1 knows p, and agent 2 does not know that if agent 1 knows p then agent 2 also knows p.

$$\neg K_1 K_2 K_1 p \wedge \neg K_2(K_1 p \Rightarrow K_2 p)$$

We will also need to express the knowledge of a group G of agents. In particular, we use E_G for everyone in G knows, C_G for common knowledge in G, and D_G for distributed knowledge in G. The syntax is then,

$$E_G\varphi,$$
$$C_G\varphi, \text{ and}$$
$$D_G\varphi$$

for any formula φ and any nonempty $G \subseteq \{1, \dots, n\}$. Moreover, we write

$$E_G^k,\ k \geq 0, \text{ for } \underbrace{E_G \cdots E_G}_{k \text{ times}}$$

Semantically,

$$\begin{array}{lll} (M, s) \models E_G\varphi & \text{iff} & (M, s) \models K_i\varphi \text{ for all } i \in G, \\ (M, s) \models C_G\varphi & \text{iff} & (M, s) \models E_G^k\varphi \text{ for } k \geq 1, \text{ and} \\ (M, s) \models D_G\varphi & \text{iff} & (M, t) \models \varphi \text{ for all } t \text{ such that } \langle s, t\rangle \in \cap_{i \in G}\mathcal{K}_i. \end{array}$$

The formula $C_G\varphi$ therefore has the meaning that everyone knows φ and everyone knows that everyone knows φ and everyone knows this ad infinitum. The formula $D_G\varphi$, in turn, has the meaning that the combined knowledge of the group is enough to know φ: the group does not consider worlds that are not possible for every agent in the group.

It is instructive to regard C_G and D_G as properties of the Kripke structure as a graph.

Definition 1.6. Call the state t G-reachable in $k > 0$ steps from s if there are states

$$s = s_0, \dots, s_k = t$$

such that for $j = 1, \dots, k$, $\langle s_{j-1}, s_j\rangle \in \mathcal{K}_i$ for any $i \in G$.

In other words, t is G-reachable from s if there is a path from s to t.

Lemma 1.7. *$(M, s) \models E_G^k\varphi$ iff $(M, t) \models \varphi$ for all t that are G-reachable from s in k steps.*

Proof. By induction on k. □

Corollary 1.8. *$(M, s) \models C_G\varphi$ iff $(M, t) \models \varphi$ for all t that are G-reachable from s.*

If the group G consists of all agents, then we leave out the G and write E, C, and D.

2 The muddy children

A group of n rational and truthful children are playing together. They all know that if they become muddy they will be punished. Along comes the father and says "At least one of you has mud on your forehead", telling them what they possibly already know. The father then asks repeatedly "Does any of you know that you have mud on your forehead?" Assuming the children answers simultaneously, what will happen?

We will assume that it is common knowledge that the father speaks the truth, that all children hear their father, that the children can see the others foreheads but not their own, and that they are rational and truthful. Let m_i, $1 \leq i \leq n$, be a primitive proposition that expresses whether child number i is muddy or not. Then, there are 2^n truth assignments over the propositions, each one corresponding to a possible world.

Let the Kripke structure $M = (S, \pi, \mathcal{K}_1, \ldots, \mathcal{K}_n)$ be a model of the problem. Namely, define

$$\begin{aligned} S &= \{\pi_1, \ldots, \pi_{2^n}\}, \\ \pi(\pi_k) &= \pi_k, \end{aligned}$$

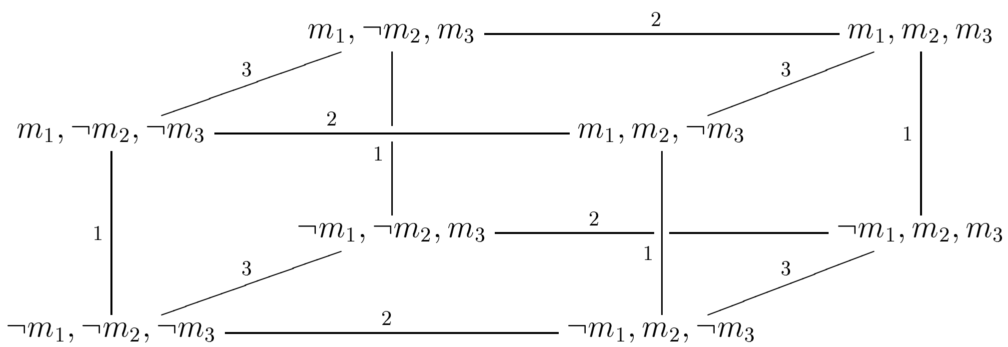
and

$$\mathcal{K}_i = \{\langle \pi_a, \pi_b \rangle \mid \pi_a(m_j) = \pi_b(m_j) \text{ for } i \neq j\}$$

when $1 \leq i, j \leq n$ and $1 \leq a, b, k \leq 2^n$, where

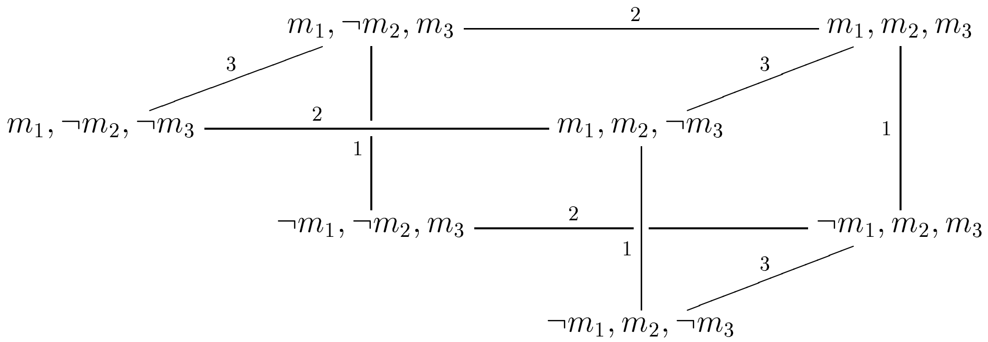
$$\pi_k(m_i) = \begin{cases} true & \text{if child } i \text{ is muddy in world } \pi_k, \\ false & \text{if not.} \end{cases}$$

Example 2.1. The case $n = 3$ (self loops and arrowheads omitted):



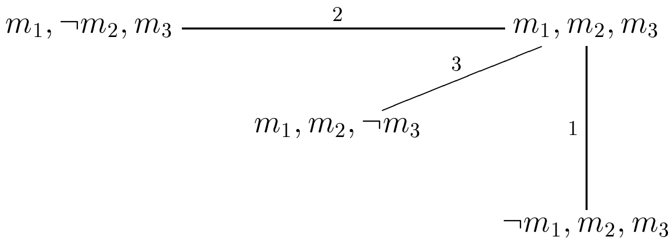
Notice that a child in a particular state can not know that it is muddy, since there is always an edge in the child's $\mathcal{K}$-relation going to a world in which it is not muddy.

After the father has said "At least one of you has mud on your forehead", no child considers the world $\neg m_1, \neg m_2, \neg m_3$ possible any more, so we have to remove the edges connected to it. In fact, we may as well remove the world too.



Now consider the state $m_1, \neg m_2, \neg m_3$. As there is no edge labeled 1 going to another state, child number 1 knows m_1, i.e., it knows that it is muddy. The same holds for child 2 in state $\neg m_1, m_2, \neg m_3$ and child 3 in state $\neg m_1, \neg m_2, m_3$. Therefore, one child will answer yes to the father's first question if there is exactly one muddy child.

If no child answers yes, then the previous three states are not possible and the Kripke structure is even smaller.



In this case, the children with muddy foreheads will answer yes to the second question if exactly two children are muddy. Finally, if no one answers yes then we are left with only one state m_1, m_2, m_3 and its self loops. Hence, all children answers yes to the third question.

Also note that if the father never utters his first statement, then we cannot remove the edges to the state $\neg m_1, \neg m_2, \neg m_3$ and a child can never be certain that it has a muddy forehead.

Proposition 2.2. *If there are k muddy children, then the muddy ones answer yes to question number k.*

Proof. By induction on k. □

3 Properties of knowledge

When a formula φ holds in all states of a Kripke structure M, we write $M \models \varphi$ and say that φ is valid in M. If φ holds in at least one state of M we say that φ is satisfiable in M. Similarly, when φ is valid in all Kripke structures we write $\models \varphi$ and say that it is valid. The formula φ is satisfiable if there is a Kripke structure in which φ is satisfiable.

Proposition 3.1. *The formula φ is valid in M iff $\neg\varphi$ is not satisfiable in M.*

Proof. If φ is valid in $M = (S, \pi, \mathcal{K}_1, \dots, \mathcal{K}_n)$, then $(M, s) \models \varphi$ for all states $s \in S$, and consequently $(M, s) \not\models \neg\varphi$ for all states $s \in S$. Hence, $\neg\varphi$ is not satisfiable in M.

On the other hand, if φ is not valid in M, then there exists a state $s \in S$ such that $(M, s) \not\models \varphi$. Thus, $(M, s) \models \neg\varphi$ and $\neg\varphi$ is satisfiable in M. $\square$

It seems reasonable to assume that if an agent knows φ and knows that $\varphi \Rightarrow \psi$, then the agent also knows ψ. This is known as the distribution axiom and is true.

Theorem 3.2. *For all $i = 1, \dots, n$,*

$$\models \big(K_i\varphi \wedge K_i(\varphi \Rightarrow \psi)\big) \Rightarrow K_i$$

Proof. Let $M = (S, \pi, \mathcal{K}_1, \dots, \mathcal{K}_n)$ be any Kripke structure. If for any $s \in S$

$$(M, s) \models K_i\varphi \wedge K_i(\varphi \Rightarrow \psi)$$

then

$$(M, t) \models \varphi \qquad \text{and} \qquad (M, t) \models \varphi \Rightarrow \psi$$

for all t such that $\langle s, t \rangle \in \mathcal{K}_i$. Hence, $(M, t) \models \psi$ and $(M, s) \models K_i\psi$. $\square$

Furthermore, one would also expect that all agents know all valid formulas. They do, and this fact is called the knowledge generalization rule.

Theorem 3.3. *For all structures $M = (S, \pi, \mathcal{K}_1, \dots, \mathcal{K}_n)$,*

$$M \models \varphi \qquad \textit{implies} \qquad M \models K_i\varphi$$

Proof. If $M \models \varphi$, then $(M, s) \models \varphi$ for all states $s \in S$. Fix s. Then, $(M, t) \models \varphi$ for all t such that $\langle s, t \rangle \in \mathcal{K}_i$. Hence, $(M, s) \models K_i\varphi$ and since s was arbitrary $M \models K_i\varphi$. $\square$

This is about all we can say about knowledge in general Kripke structures. However, by restricting the class of possibility relations we will find new interesting properties.

In the class of reflexive Kripke structures anything an agent knows is true. This is called the knowledge axiom and is what distinguishes knowledge from belief. If you believe something, it is not necessarily true. But if you know it, it is true.

A Kripke structure $M = (S, \pi, \mathcal{K}_1, \dots, \mathcal{K}_n)$ is reflexive if $\langle s, s\rangle \in \mathcal{K}_i$ for all $s \in S$.

Theorem 3.4. *For all reflexive structures $M = (S, \pi, \mathcal{K}_1, \dots, \mathcal{K}_n)$,*

$$M \models K_i\varphi \Rightarrow \varphi$$

Proof. If $(M, s) \models K_i\varphi$, then $(M, t) \models \varphi$ for all t such that $\langle s, t\rangle \in \mathcal{K}_i$. Since $\mathcal{K}_i$ is reflexive, $\langle s, s\rangle \in \mathcal{K}_i$ and $(M, s) \models \varphi$. □

If whenever an agent knows φ it knows that it knows φ, then it is capable of positive introspection. The positive introspection axiom holds in the transitive Kripke structures.

A relation $\mathcal{K}_i$ is transitive when for all $s, t, u \in S$,

$$\langle s, t\rangle \in \mathcal{K}_i \quad \text{and} \quad \langle t, u\rangle \in \mathcal{K}_i \quad \text{implies} \quad \langle s, u\rangle \in \mathcal{K}_i$$

Theorem 3.5. *For all transitive structures $M = (S, \pi, \mathcal{K}_1, \dots, \mathcal{K}_n)$,*

$$M \models K_i\varphi \Rightarrow K_iK_i\varphi$$

Proof. Assume $(M, s) \models K_i\varphi$, and consider any t for which $\langle s, t\rangle \in \mathcal{K}_i$. For all u such that $\langle t, u\rangle \in \mathcal{K}_i$, $\langle s, u\rangle \in \mathcal{K}_i$ by transitivity. Hence, $(M, u) \models \varphi$ and $(M, t) \models K_i\varphi$. Consequently, $(M, s) \models K_iK_i\varphi$. □

Finally, an agent is capable of negative introspection in the symmetric and transitive Kripke structures. That is, if the agent does not know φ, then it knows that it does not know φ.

A relation $\mathcal{K}_i$ is symmetric if for all $s, t \in S$,

$$\langle s, t\rangle \in \mathcal{K}_i \quad \text{implies} \quad \langle t, s\rangle \in \mathcal{K}_i$$

Theorem 3.6. *For all symmetric and transitive Kripke structures $M = (S, \pi, \mathcal{K}_1, \dots, \mathcal{K}_n)$,*

$$M \models \neg K_i\varphi \Rightarrow K_i\neg K_i\varphi$$

Proof. Suppose $(M, s) \models \neg K_i\varphi$. Then, there is a state t such that $\langle s, t\rangle \in \mathcal{K}_i$ and $(M, t) \not\models \varphi$. Take any u for which $\langle s, u\rangle \in \mathcal{K}_i$. By symmetry $\langle u, s\rangle \in \mathcal{K}_i$ and by transitivity $\langle u, t\rangle \in \mathcal{K}_i$. Thus,

$$(M, u) \models \neg K_i\varphi \qquad \text{and} \qquad (M, s) \models K_i\neg K_i\varphi$$

□

We now turn to the properties of E_G and C_G. Since $E_G\varphi$ is true when all agents know φ, we arrive at the following characterization.

Theorem 3.7. *For all structures $M = (S, \pi, \mathcal{K}_1, \ldots, \mathcal{K}_n)$ and all nonempty $G \subseteq \{1, \ldots, n\}$,*

$$M \models E_G\varphi \Leftrightarrow \bigwedge_{i\in G} K_i\varphi$$

Proof. Follows immediately from the definition of E_G. □

A formula φ is common knowledge if everyone knows φ and everyone knows this recursively. The fixed-point axiom is a formalization of this.

Theorem 3.8. *For all structures $M = (S, \pi, \mathcal{K}_1, \ldots, \mathcal{K}_n)$ and all nonempty $G \subseteq \{1, \ldots, n\}$,*

$$M \models C_G\varphi \Leftrightarrow E_G(\varphi \wedge C_G\varphi)$$

Proof. Assume $(M, s) \models C_G\varphi$. Then, $(M, t) \models \varphi$ for all states t that are G-reachable from s. Pick a state u such that $\langle s, u\rangle \in \mathcal{K}_i$ for some $i \in G$. Since every state G-reachable from u is also G-reachable from s,

$$(M, u) \models \varphi \wedge C_G\varphi$$

Hence, $(M, s) \models E_G(\varphi \wedge C_G\varphi)$.

Conversely, assume $(M, s) \models E_G(\varphi \wedge C_G\varphi)$. Then, $(M, t) \models \varphi \wedge C_G\varphi$ for all t such that $\langle s, t\rangle \in \mathcal{K}_i$ for some $i \in G$. Take any u G-reachable from s. If u is G-reachable from s in one step, then $(M, u) \models \varphi$, and if u is G-reachable from s in k steps, then there is a state t G-reachable in one step from s such that u is G-reachable in $k - 1$ steps from t. Hence, $(M, t) \models C_G\varphi$ and $(M, u) \models \varphi$. In conclusion, $(M, s) \models C_G\varphi$. □

Finally, the induction rule presents us with another property of common knowledge.

Theorem 3.9. *For all structures $M = (S, \pi, \mathcal{K}_1, \dots, \mathcal{K}_n)$ and all nonempty $G \subseteq \{1, \dots, n\}$,*

$$M \models \varphi \Rightarrow E_G(\varphi \wedge \psi) \qquad \textit{implies} \qquad M \models \varphi \Rightarrow C_G\psi$$

Proof. Suppose $M \models \varphi \Rightarrow E_G(\varphi \wedge \psi)$ and $(M, s) \models \varphi$. Then, $(M, s) \models E_G(\varphi \wedge \psi)$ and $(M, t) \models \varphi \wedge \psi$ for all t G-reachable from s in one step.

Assume $(M, t) \models \varphi \wedge \psi$ holds for all t G-reachable from s in less than k steps and let u be G-reachable from s in k steps. Then, there exists a t G-reachable from s in $k - 1$ steps such that u is G-reachable from t in one step. Hence, $(M, t) \models \varphi \wedge \psi$ and $(M, t) \models \varphi \Rightarrow E_G(\varphi \wedge \psi)$ from which it follows that $(M, u) \models \varphi \wedge \psi$.

Thus, by induction, $(M, u) \models \varphi \wedge \psi$ for all u G-reachable from s. Hence, $(M, s) \models C_G\psi$. □