Epistemic logics: an introduction

Valentin Goranko
Technical University of Denmark
DTU Informatics

November 2010

Modal reasoning about knowledge and belief

- Epistemic reading of the modal operators:

  $\Box\varphi$: 'the agent knows that φ';

  $\Diamond\varphi$: 'φ is consistent with the agent's knowledge'.

- Doxastic reading of the modal operators:

  $\Box\varphi$: 'the agent believes that φ';

  $\Diamond\varphi$: 'φ is consistent with the agent's beliefs'.

- Knowledge is always true, while beliefs need not be.

Epistemic logic: syntax and basic principles

Language of EL: just like the basic modal logic, but with the knowledge operator K instead of $\Box$. Formulae:

$$\varphi = p \mid \bot \mid \neg\varphi \mid \varphi \wedge \varphi \mid \mathrm{K}\varphi$$

The other propositional connectives: definable as usual.

No special notation for the dual of K.

Some basic principles of EL:

K $\mathrm{K}(\phi \to \psi) \to (\mathrm{K}\phi \to \mathrm{K}\psi)$

T $\mathrm{K}\varphi \to \varphi$ (knowledge is truthful)

4 $\mathrm{K}\varphi \to \mathrm{KK}\varphi$ (positive introspection)

5 $\neg\mathrm{K}\varphi \to \mathrm{K}\neg\mathrm{K}\varphi$ (negative introspection)

Thus, EL is in fact the modal logic of equivalence relations S5.

Problem: logical omniscience.

Kripke models for the epistemic logic

- Epistemic frame: a pair (W, R), where:
  - W is a non-empty set of possible worlds, representing the possible states of affairs in the actual world.
  - $R \subseteq W^2$ is an equivalence relation, called epistemic indistinguishability relation between possible worlds.
- epistemic model: $\mathcal{M} = (W, R, V)$ where (W, R) is an epistemic frame and $V : \mathrm{AP} \to \mathcal{P}(W)$ is a valuation assigning to every atomic proposition the set of possible worlds where it is true.

The idea of the epistemic indistinguishability relation:

$$s_1 R s_2$$

holds if, from all that the agent knows, he cannot distinguish the states s_1 and s_2. In other words, at the state s_1 the agent considers s_2 equally possible to be the case.

Kripke semantics of the epistemic logic

The semantics for EL is the usual Kripke semantics. In particular:

$\mathcal{M}, s \models \mathrm{K}\varphi$ iff $\mathcal{M}, t \models \varphi$ for every state t such that sRt

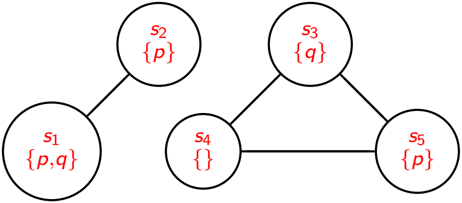
Meaning: the agent knows φ at the possible world s if φ is true at every possible world t that is indistinguishable from s by the agent.

That is, the agent knows φ at the possible world s if (s)he has no uncertainty about the truth of φ at that world.

Epistemic models: example 1

Consider a language with two atomic propositions, p and q.

Consider the model $\mathcal{M}$ (the reflexive loops are omitted):



- $\mathcal{M}, s_1 \models p \wedge \mathrm{K}p$; $\mathcal{M}, s_1 \models q \wedge \neg\mathrm{K}q$; $\mathcal{M}, s_1 \models \mathrm{K}\mathrm{K}p \wedge \mathrm{K}\neg\mathrm{K}q$.
- $\mathcal{M}, s_3 \models q \wedge \neg p \wedge \neg\mathrm{K}q \wedge \neg\mathrm{K}\neg p \wedge \mathrm{K}(\neg\mathrm{K}q \wedge \neg\mathrm{K}\neg p)$.

Epistemic models: example 2

See Pacuit's slides.

Multi-agent epistemic reasoning: a prelude

Suppose now that there are two agents, Ann and Bob.
We associate knowledge operators with each of them.

- $\mathrm{K}_A p$: "Ann knows that p".
- $\mathrm{K}_B p$: "Bob knows that p".
- $\mathrm{K}_A \mathrm{K}_B p$: "Ann knows that Bob knows that p".
- $\mathrm{K}_{AB} p := \mathrm{K}_A p \wedge \mathrm{K}_B p$: "Both Ann and Bob know that p".
- There can be many agents.
  So, let $\mathrm{E}p$ mean "Everybody knows that p".
- Then $\mathrm{EE}p$: "Everybody knows that everybody knows that p".
- $\mathrm{EEE} \ldots p$ mean "Everybody knows that everybody knows that that everybody knows ... that p".

  That means "p is a common knowledge".

Multi-agent epistemic operators

Framework: a set of agents (players) $\mathbb{A}\mathbf{g}$, each possessing certain knowledge about the system, the environment, themselves, and the other agents.

Multi-agent epistemic logics: multi-modal logics with *epistemic modalities* for agents and groups (*coalitions*) of agents.

- $\mathrm{K}_{\mathsf{i}}\varphi$: *'The agent* i *knows that* φ*'.*
- $\mathrm{K}_{\mathbf{A}}\varphi$: *'Every agent in the group* **A** *knows that* φ*'.*
- $\mathrm{D}_{\mathbf{A}}\varphi$: *'It is a distributed knowledge amongst the agents in the group* **A** *implies that* φ*'.*

  or, *'The collective knowledge of all agents in the group* **A** *implies that* φ*'.*
- $\mathrm{C}_{\mathbf{A}}\varphi$: *'It is a common knowledge amongst the agents in the group* **A** *that* φ*'.*

Multi-agent epistemic operators: distributed knowledge

The idea: if Agent 1 knows φ and Agent 2 knows ψ, then they together can derive $\varphi \wedge \psi$, i.e. $\mathrm{D}_{1,2}\, \varphi \wedge \psi$ holds.

Important concept in distributed computing.

Multi-agent epistemic operators: common knowledge

Intuitively, φ is a common knowledge amongst the agents in **A** if $\mathrm{K}_{\mathbf{A}}\varphi$ holds, and $\mathrm{K}_{\mathbf{A}}\mathrm{K}_{\mathbf{A}}\varphi$ holds, and $\mathrm{K}_{\mathbf{A}}\mathrm{K}_{\mathbf{A}}\mathrm{K}_{\mathbf{A}}\varphi$ holds, etc. – infinitely!

This cannot be reduced to a finite chain.

Example: the coordinated attack problem.

The coordinated attack problem

- Two allied armies are on the two sides of a mountain, and their common enemy is in a fortress on top of the mountain.
- Neither army can defeat the enemy alone, and both army commanders know that. So, they have to attack together.
- There are two options for the simultaneous attack: at down or at night.
- The two commanders must coordinate the time of the attack, by confirming their choice between themselves.
- That is, the time of the attack must become their common knowledge!
- Their only means for communication is by sending messengers to each other.
- Can the attack be coordinated reliably?

The muddy children problem

See Pacuit's slides.

Multi-agent epistemic logic (MAEL): formal syntax

Formal syntax:

$$\varphi := \boldsymbol{p} \mid \neg\varphi \mid \varphi \vee \psi \mid \mathrm{K}_{\mathbf{i}}\varphi \mid \mathrm{K}_{\mathbf{A}}\varphi \mid \mathrm{C}_{\mathbf{A}}\varphi \mid \mathrm{D}_{\mathbf{A}}\varphi,$$

where **i** is an agent, **A** is an arbitrary set of agents, and $\mathrm{K}_{\mathrm{i}}, \mathrm{K}_{\mathbf{A}}, \mathrm{C}_{\mathbf{A}}, \mathrm{D}_{\mathbf{A}}$ are epistemic modal operators respectively for individual, group, common, and distributed knowledge of agents and coalitions.

Multi-agent epistemic logic: expressing some epistemic properties

- Compare: $\neg \mathrm{K}_1 \varphi$ and $\mathrm{K}_1 \neg \varphi$
- "Agent 1 does not know whether φ is true:" $\neg \mathrm{K}_1 \varphi \wedge \neg \mathrm{K}_1 \neg \varphi$
- "The knowledge of agent 1 about φ is consistent:"
$$\mathrm{K}_1 \varphi \rightarrow \neg \mathrm{K}_1 \neg \varphi$$
- "Agent 2 knows that agent 1 does not know whether φ is true:"
$$\mathrm{K}_2(\neg \mathrm{K}_1 \varphi \wedge \neg \mathrm{K}_1 \neg \varphi)$$
- $\mathrm{K}_{\{1,2\}} \varphi \wedge \mathrm{K}_{\{1,2\}} \mathrm{K}_{\{1,2\}} \varphi \wedge \neg \mathrm{C}_{\{1,2\}} \varphi$

"Both agents 1 and 2 know that φ is true, and they both know that they both know it, but the truth of φ is not a common knowledge between them".

- $\neg \mathrm{K}_1 \varphi \wedge \neg \mathrm{K}_2 \varphi \wedge \mathrm{C}_{\{1,2\}}(\neg \mathrm{K}_1 \varphi \wedge \neg \mathrm{K}_2 \varphi) \wedge \mathrm{D}_{\{1,2\}} \varphi$

"None of the agents 1 and 2 knows that φ is true, and that is a common knowledge between them, but the truth of φ is distributed knowledge between them".

Using Multi-agent epistemic logic: the 3 cards scenario

There are 3 cards: A(ce), K(ing) and Q(ueen) and three persons: 1,2,3. Each of them holds one of the cards and does not know the cards of the other two.

To describe the situation in MAEL, we introduce propositions: $P_{i,A}, P_{i,K}, P_{i,Q}$, for $i = 1, 2, 3$, where $P_{i,A}$ means that the person i holds the card A, etc. Here are some true formulae:

- $P_{1,A} \lor P_{2,A} \lor P_{3,A}$; $P_{1,A} \lor P_{1,K} \lor P_{1,Q}$;
- $\mathrm{K}_{\{1,2,3\}}(P_{1,A} \lor P_{2,A} \lor P_{3,A})$; $\mathrm{K}_{\{1,2,3\}}(P_{1,A} \lor P_{1,K} \lor P_{1,Q})$;
- $\mathrm{C}_{\{1,2,3\}}(P_{1,A} \lor P_{2,A} \lor P_{3,A})$; $\mathrm{C}_{\{1,2,3\}}(P_{1,A} \lor P_{1,K} \lor P_{1,Q})$;
- $P_{i,A} \to \mathrm{K}_i P_{i,A}$; $\mathrm{C}_{\{1,2,3\}}(P_{i,A} \to \mathrm{K}_i P_{i,A})$; $\neg P_{i,A} \to \mathrm{K}_i \neg P_{i,A}$;
- $P_{1,A} \to \mathrm{K}_1 \neg P_{2,A} \land \neg \mathrm{K}_1 P_{2,K} \land \neg \mathrm{K}_1 P_{2,Q}$;
- $\mathrm{D}_{\{1,2\}} P_{3,A} \lor \mathrm{D}_{\{1,2\}} P_{3,K} \lor \mathrm{D}_{\{1,2\}} P_{3,Q}$;
- $\mathrm{C}_{\{1,2,3\}}(P_{1,A} \land P_{2,K} \to \mathrm{D}_{\{1,2\}} P_{3,Q})$.

Multi-agent epistemic logic: Kripke models

Multi-agent epistemic model:

$$\mathcal{M} = \langle \mathbf{S}, \Pi, \pi, \mathbb{A}\mathbf{g}, \sim_{\mathbf{1}}, ..., \sim_{\mathbf{n}} \rangle \,,$$

where:

- $\mathbf{S}$ is a set of states,
- Π is a set of atomic propositions,
- $\pi : \mathbf{S} \to \mathbf{2}^{\Pi}$ is a valuation,
- $\mathbb{A}\mathbf{g} = \{\mathbf{1}, ..., \mathbf{n}\}$ is a finite set of agents,
- $\sim_{\mathbf{1}}, ..., \sim_{\mathbf{n}}$ – the epistemic indistinguishability relations associated with the agents.

Multi-agent epistemic logic: formal semantics

The formal semantics of the epistemic operators at a state in a multi-agent epistemic model

$$\mathcal{M} = \langle \mathbf{S}, \Pi, \pi, \mathbb{A}\mathbf{g}, \sim_1, ..., \sim_n \rangle$$

is given by the clauses:

$(\mathrm{K_i})$ $\mathcal{M}, q \models \mathrm{K_i}\varphi$ iff $\mathcal{M}, q' \models \varphi$ for all q' such that $q \sim_i q'$.

$(\mathrm{K_A})$ $\mathcal{M}, q \models \mathrm{K_A}\varphi$ iff $\mathcal{M}, q' \models \varphi$ for all q' such that $q \sim_A^E q'$, where $\sim_A^E = \bigcup_{i \in A} \sim_i$.

$(\mathrm{C_A})$ $\mathcal{M}, q \models \mathrm{C_A}\varphi$ iff $\mathcal{M}, q' \models \varphi$ for all q' such that $q \sim_A^C q'$, where $\sim_A^C$ is the transitive closure of $\sim_A^E$.

$(\mathrm{D_A})$ $\mathcal{M}, q \models \mathrm{D_A}\varphi$ iff $\mathcal{M}, q' \models \varphi$ for all q' such that $q \sim_A^D q'$, where $\sim_A^D = \bigcap_{i \in A} \sim_i$.

Epistemic model updates

- Epistemic models represent the static knowledge of the agents at a given moment.
- When the knowledge of any agent changes, the model must be updated to reflect that change.
- These updates are studied by Dynamic epistemic logic.
- For instance, the knowledge of agents changes as a result of communication.
- A simplest form of communication is public announcement. It creates a common knowledge amongst all agents of the truth of the publicly announced fact.
- The model update after public announcement of the truth of φ is simple: remove all states where φ is false.

Modeling and solving the muddy children problem

See Pacuit's slides.

Axioms for the basic multi-agent epistemic logic $\mathbf{S5}_n$

The multi-modal logic $\mathbf{S5}_n$ is a multi-modal version of **S5**.

The axioms of $\mathbf{S5}_n$:

K $\mathrm{K_i}(\phi \to \psi) \to (\mathrm{K_i}\phi \to \mathrm{K_i}\psi)$

T $\mathrm{K_i}\varphi \to \varphi$ (knowledge is truthful)

4 $\mathrm{K_i}\varphi \to \mathrm{K_i}\mathrm{K_i}\varphi$ (positive introspection)

5 $\neg\mathrm{K_i}\varphi \to \mathrm{K_i}\neg\mathrm{K_i}\varphi$ (negative introspection)

Inference rules: for each i:

$$\frac{\varphi}{\mathrm{K_i}\varphi}.$$

Axioms for common and distributed knowledge

MAEL$_n$ extends **S5**$_n$ with the following schemes for each $\mathsf{A} \subseteq \mathbb{A}\mathbf{g}$:

- The axioms for K_A: $\mathrm{K}_\mathsf{A}\varphi \leftrightarrow \bigwedge_{i\in\mathsf{A}} \mathrm{K}_i\varphi$;
- (Least) fixed point axioms for C_A:

$$\mathbf{LFP}_\mathsf{A}: \quad \mathrm{C}_\mathsf{A}\varphi \leftrightarrow (\varphi \land \mathrm{K}_\mathsf{A}\mathrm{C}_\mathsf{A}\varphi),$$

- Axioms for D_A:

$$\begin{array}{ll} \mathbf{S5}(\mathrm{D}_\mathsf{A}): & \text{The } S5 \text{ axioms for } \mathrm{D}_\mathsf{A}, \\ \mathrm{D}_\mathbf{i}: & \mathrm{D}_i\varphi \leftrightarrow \mathrm{K}_i\varphi, \\ \mathbf{INCL}(\mathrm{D}): & \mathrm{D}_\mathsf{A}\varphi \to \mathrm{D}_\mathsf{B}\varphi \text{ whenever } \mathsf{A} \subseteq \mathsf{B}. \end{array}$$

and, for each $i \in \mathbb{A}\mathbf{g}$, the inference rule

$$\frac{\varphi}{\mathrm{K}_i\varphi.}$$

Exercise: derive Segerberg's induction axiom:

$$\mathbf{IND}_\mathsf{A}: \varphi \land \mathrm{C}_\mathsf{A}(\varphi \to \mathrm{K}_\mathsf{A}\varphi) \to \mathrm{C}_\mathsf{A}\varphi;$$

Multi-agent Epistemic Logic: some technical results

- The axiom system of the KC-fragment of $\mathbf{MAEL}_n$ is complete. Satisfiability/validity in $\mathbf{MAEL}_n^{\mathrm{KC}}$ is decidable.

Adding common knowledge, however, increases the complexity, from PSPACE-complete for $\mathbf{S5}_n$ to EXPTIME-complete.

- The axiom system $\mathbf{MAEL}_n^{\mathrm{KD}}$ of the KD-fragment of **MAEL** is sound and complete. $\mathbf{MAEL}_n^{\mathrm{KD}}$ is decidable, EXPTIME-complete.

- The full axiom system $\mathbf{MAEL}_n$ is sound and complete. $\mathbf{MAEL}_n$ is decidable, EXPTIME-complete.

Tableau based decision procedures for multi-agent epistemic logics

- V. Goranko and D. Shkatov:
Tableau-based decision procedure for the multi-agent epistemic logic with operators of common and distributed knowledge: http://arxiv.org/abs/0808.4133

- Recently implemented by Thomas Vestergaard and available online:

  http://www.thomaslyngby.dk/thesis/

- V. Goranko and D. Shkatov. Tableau-based Procedure for Deciding Satisfiability in the Full Coalitional Multi-agent Epistemic Logic: http://arxiv.org/abs/0902.2125

  Waiting for implementation!

Exercises on multi-agent epistemic logics: the three cards scenario revisited

♦ Construct a Kripke model describing the three cards scenario. Take as states all possible distributions of the cards.

♦ Construct the update of that model after person 1 privately tells person 2 that he holds the Ace.

♦ Find epistemic formulae, that are false in the original model but become true in the updated one, after the announcement of 1.

♦ Using the updated model, argue that, after the announcement of person 1, person 3 still does not know the card of person 2.