

Epistemic Logic

Natasha Alechina Brian Logan

Utrecht University

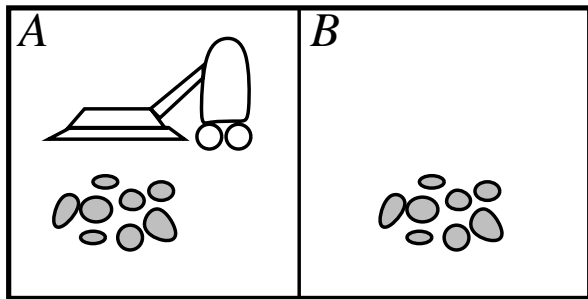
n.a.alechina@uu.nl b.s.logan@uu.nl

Why imperfect information

- agents can only select actions based on the information they have
- different agents have different information
- representing this is important for communication, coordinating actions, or keeping opponents in the dark
- just as temporal logic describes the system dynamics . . .
- epistemic logic describes the agents' knowledge

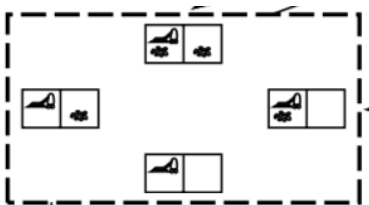
Motivating example: representing imperfect information

Recall the vacuum cleaner the from Russell and Norvig's textbook:



Motivating example: representing imperfect information

- suppose the agent has a location sensor but no dirt sensor
- then it will 'know' that it is in room A, but will consider several states of the world possible regarding the dirt:



Possible worlds

- this is a key idea behind representing incomplete information: the agent considers several states of the world **possible**
- or (equivalently) it cannot distinguish between several possible worlds
- this **indistinguishability relation** is the main idea behind both epistemic logic and in representing agent systems where agents may have imperfect information

Epistemic logic

- Epistemic Logic (EL) is a logic for reasoning about knowledge
- its models are **Kripke models**
- they look very much like state transition systems, but instead of transition relation they have an indistinguishability relation (for every agent in the system)

Kripke Models

Definition (Kripke model)

Let \mathcal{PV} be a set of atomic propositions (p, q, r, \dots) and Agt a finite set of agents.

A **Kripke model** $\mathcal{M} = \langle St, \{\sim_i \mid (i \in Agt)\}, \mathcal{V} \rangle$ consists of

- a non-empty set of states St
- an indistinguishability relation \sim_i for each agent i (which is reflexive, transitive and symmetric)
- a **valuation of propositions** $\mathcal{V} : \mathcal{PV} \rightarrow 2^{St}$

Language of epistemic logic

Definition (Syntax of EL)

$$\varphi := p \mid \neg\varphi \mid \varphi \wedge \psi \mid K_i\varphi$$

- where $p \in \mathcal{PV}$ and $i \in \mathcal{Agt}$
- sometimes useful to define $\bar{K}_i\varphi := \neg K_i\neg\varphi$ (for i considers φ possible)

Possible World Semantics

Definition (Semantic Clauses)

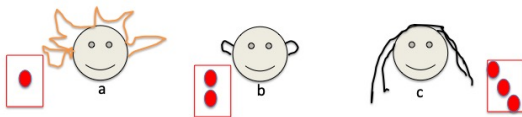
- $\mathcal{M}, q \models p$ iff $q \in \mathcal{V}(p)$;
- $\mathcal{M}, q \models \neg\varphi$ iff not $\mathcal{M}, q \models \varphi$;
- $\mathcal{M}, q \models \varphi \wedge \psi$ iff $\mathcal{M}, q \models \varphi$ and $\mathcal{M}, q \models \psi$;
- $\mathcal{M}, q \models K_i\varphi$ iff, for every $q' \in St$ such that $q \sim_i q'$, we have $\mathcal{M}, q' \models \varphi$.

Epistemic logic

- $K_i\varphi$ means that i knows that φ is true
- $\mathcal{M}, q \models K_i\varphi$ iff φ holds in all worlds that for the agent i are indistinguishable from q

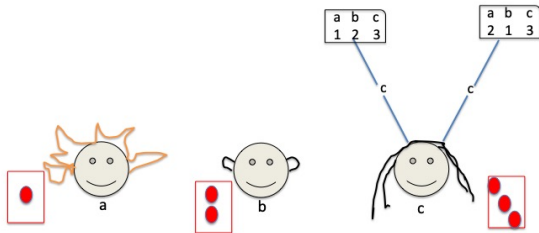
Example: agents playing cards

Propositions: $a1$ for 'agent a has a card with 1 dot', $b1, c1$, $a2$, etc.



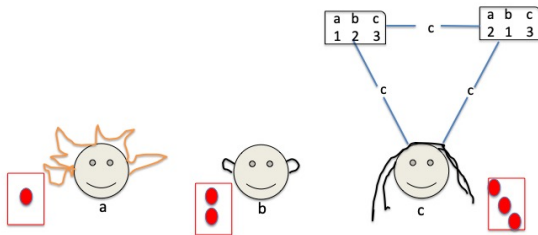
Example: agents playing cards

c considers two states possible



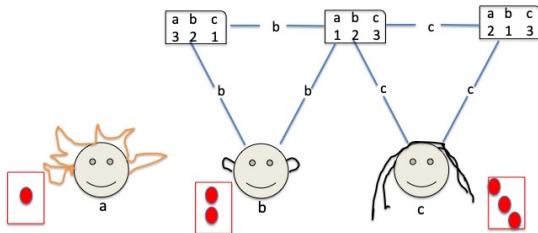
Example: agents playing cards

$K_c c3, \neg K_c b2$



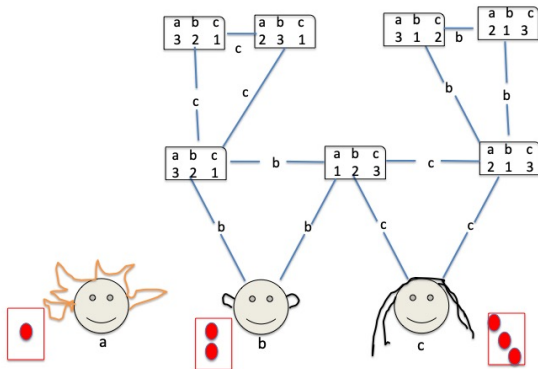
Example: agents playing cards

b is similar



Example: agents playing cards

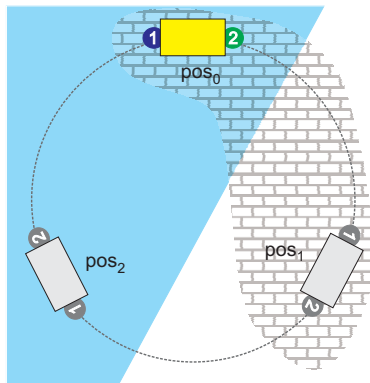
Reasoning about each other's knowledge: $K_c(K_b b1 \vee K_b b2)$,
 $K_c \neg K_b(c3)$



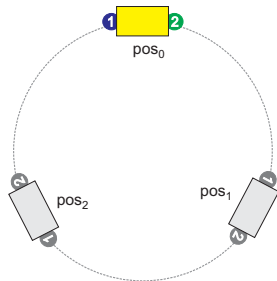
The Kripke model is not complete!

Example: Robots and Carriage

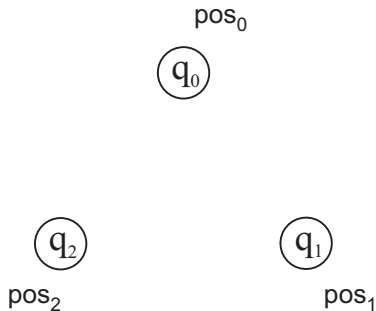
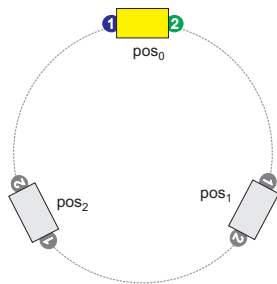
Suppose robot 1 perceives only the colour of the surface, robot 2 only the texture



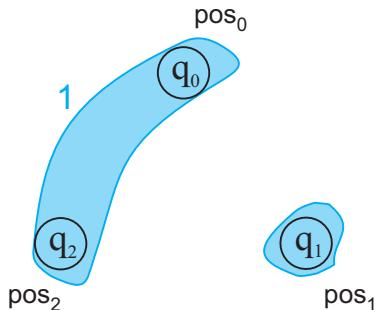
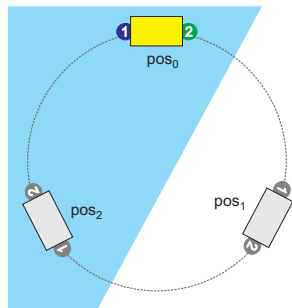
How do we model this as a Kripke model?



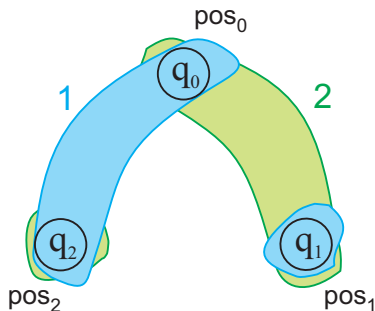
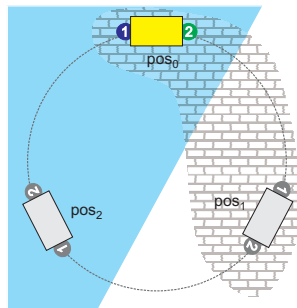
How do we model this as a Kripke model?



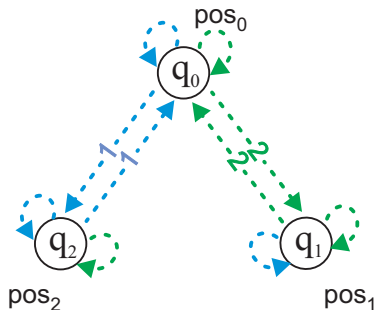
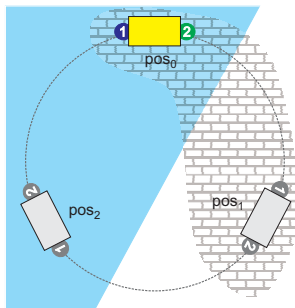
How do we model this as a Kripke model?



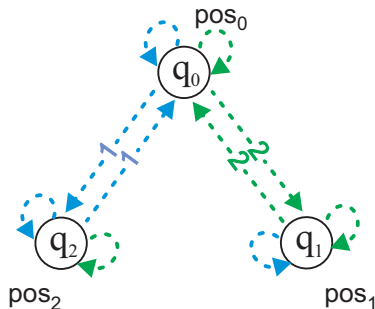
How do we model this as a Kripke model?



How do we model this as a Kripke model?



Some formulas



$$\begin{aligned} pos_2 &\rightarrow \neg K_1 pos_2 \\ pos_2 &\rightarrow K_1 \neg pos_1 \\ pos_2 &\rightarrow K_2 K_1 \neg pos_1 \end{aligned}$$

Some properties of K_i

- **knowledge is veridical**: for every φ and i , $K_i\varphi \rightarrow \varphi$ is valid, that is, true in all states in all models
- this is because \sim_i is reflexive: since $q \sim_i q$, and φ is required to be true in all states related to q by \sim_i , it is true in q
- **knowledge is consistent** $\neg K_i\perp$ is valid (for the same reason)
- **knowledge is introspective** $K_i\varphi \rightarrow K_iK_i\varphi$ is valid (this is because \sim_i is transitive) and $\neg K_i\varphi \rightarrow K_i\neg K_i\varphi$ (because \sim_i is symmetrical)
- aside: **correspondence theory** between properties of modalities (like knowledge) and first order properties of Kripke models is a fascinating subject (Johan van Benthem)

Logical Omniscience

- a less appealing property of K_i is **logical omniscience**: each agent knows all tautologies and all logical consequences of its knowledge
- $K_i \top$ is valid (because \top is true in all states reachable by \sim_i)
- $K_i(\varphi \rightarrow \psi) \rightarrow (K_i\varphi \rightarrow K_i\psi)$ is valid (if $\varphi \rightarrow \psi$ and φ are true in all states reachable by \sim_i , then ψ has to be true as well)
- if an agent is a vacuum cleaner (no logical reasoning about tautologies) this is not a problem
- but it means epistemic logic is not very suitable for modelling human-like resource-bounded **reasoners**

see, Alechina and Logan [Ascribing beliefs to resource bounded agents](#). AAMAS 2002: 881–888

Applications of epistemic logic

- analysing **distributed systems** (in the next lecture)
- verifying protocols where knowledge of participants is important (for example cryptographic protocols)
- **epistemic planning**: how to plan a sequence of actions to achieve some epistemic state

see, Bolander and Andersen [Epistemic planning for single- and multi-agent systems](#). Journal of Applied Non-Classical Logics, 21(1), 9–34