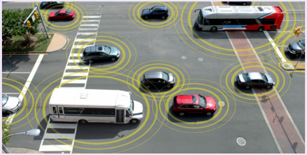
Epistemic Reasoning in Multi-agent Systems

Tristan Charrier François Schwarzentruber

École Normale Supérieure Rennes

May 14th, 2019

Automation of complex tasks



Autonomous cars

Intelligent farming

Nuclear decommissioning

cars, robots, humans

Several *agents* that interact with the environment and with each other.

Imperfect information

- Agents have local view of the environment
- Agents communicate
- Agents act

Decisions are taken with respect to *knowledge*.

Interaction relies on knowledge

if I know it is safe **then**
`I go`

if I know you are at the market place **then**
`I join you`

if (I know it is safe) and (I know you do not know it is safe) **then**
`I tell you it is safe`

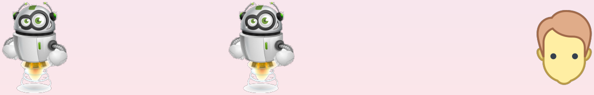
if I know you know it is safe **then**
`I do not tell you it is safe`

if I know you know I know it is safe or not **then**
`I do not wait for a message from you`

Need to build understandable multi-agent systems

Motivation

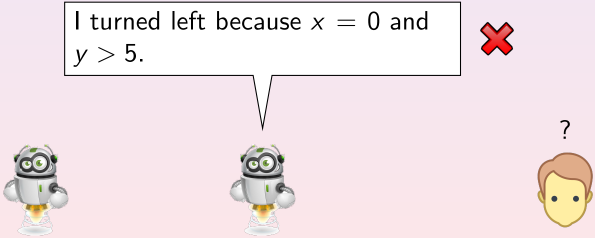
- Robots interacting with humans
- Legal issues in case of failure



Need to build understandable multi-agent systems

Motivation

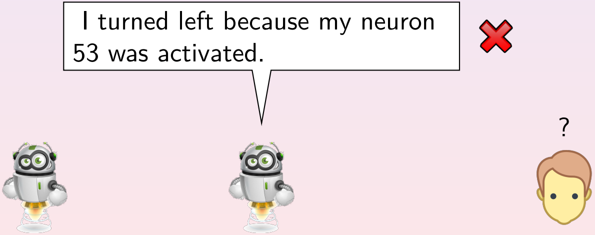
- Robots interacting with humans
- Legal issues in case of failure



Need to build understandable multi-agent systems

Motivation

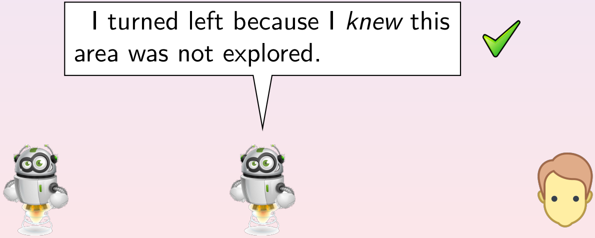
- Robots interacting with humans
- Legal issues in case of failure



Need to build understandable multi-agent systems

Motivation

- Robots interacting with humans
- Legal issues in case of failure



Solution: reasoning about knowledge

Given:

- what agents sense;
- the actions and communications that occurred

What does each agent know?

Content of this tutorial

1. Introduction to epistemic logic
   [van Ditmarsch, Joseph Y. Halpern, van der Hoek, Kooi, Chap. 1. of Handbook of epistemic logic, 2015]
2. Knowing and seeing
   [Balbiani, et al. Agents that see each other IGPL 2012]
3. Knowledge and time
   [Dixon, Nalon, Ramanujam, Chap. 5. of Handbook of epistemic logic, 2015]
4. Dynamic epistemic logic
   [Moss, Chap. 6. of Handbook of epistemic logic, 2015]
5. Knowledge-based programs
   [Joseph Y. Halpern, Moshe Vardi, Ronald Fagin et Yoram Moses. Reasoning about knowledge 1995]
   [Saffidine, Zanuttini, et al., AAAI 2018]

References

[Jaakko Hintikka. Knowledge and Belief: An Introduction to the Logic of the Two Notions (1962)]

[J-J Ch. Meyer, van der Hoek, Epistemic logic in AI and computer science, 1995]

[Joseph Y. Halpern, Moshe Vardi, Ronald Fagin et Yoram Moses. Reasoning about knowledge 1995]

[van Ditmarsch, van der Hoek, Kooi, Dynamic epistemic logic, 2007]

[van Ditmarsch, Joseph Y. Halpern, van der Hoek, Kooi, Handbook of epistemic logic, 2015]

Acknowledgment

- Tristan Charrier, former PhD student in Rennes, for many results that will be presented
- Sébastien Gamblin and Alexandre Niveau, for the implementation of succinct/symbolic models in Hintikka's World
- Sophie Pinchinat, head of the LogicA group in Rennes
- Many other colleagues: Valentin Goranko, Andreas Herzig, Emiliano Lorini, Arthur Queffelec, Abdallah Saffidine, Bruno Zanuttini, etc.

Outline

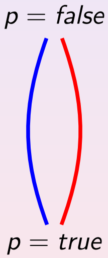
Outline

Outline

1. The Hintikka's World project
   - Motivation 1: face the difficulties in explaining possible worlds
   - Motivation 2: disseminating in many communities
   - Open software
2. Epistemic logic
3. Model checking
4. Theorem proving
5. Language properties

Once upon a time... In 2011-2012...

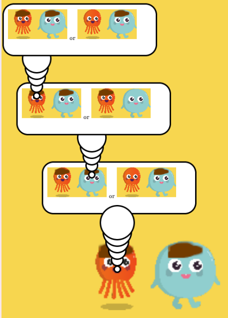
I explained epistemic logic to other researchers in logic/AI/verification...



... but nobody understood me...

Possible worlds

... but, since 2017, everybody understood me with comics...



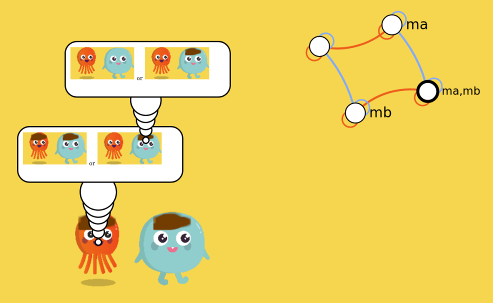
http://hintikkasworld.irisa.fr/

[demo IJCAI-ECAI 2018]

Semantics of knowing something

Agent *a* knows that *b* is dirty.

Epistemic states = pointed Kripke structures

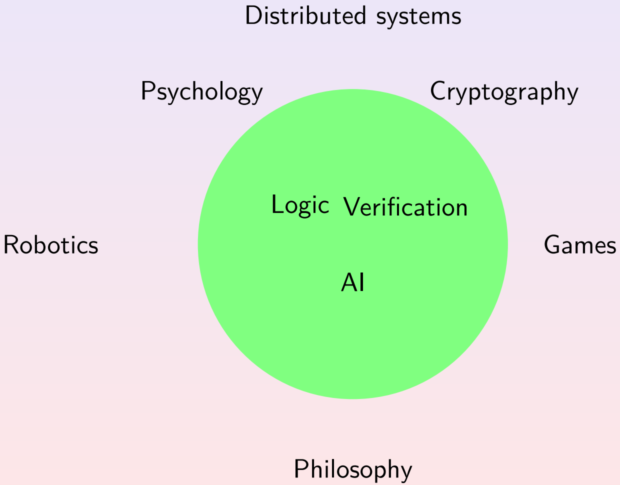


Comics = unraveling of a pointed Kripke structure.

Outline

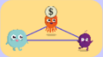
1. The Hintikka's World project
   - Motivation 1: face the difficulties in explaining possible worlds
   - Motivation 2: disseminating in many communities
   - Open software
2. Epistemic logic
3. Model checking
4. Theorem proving
5. Language properties

Explaining these models in many communities



Explaining these models in many communities

Distributed systems



Dining cryptographers problem


Sally and Ann

Psychology

Cryptography

Logic Verification

Robotics

Games

AI

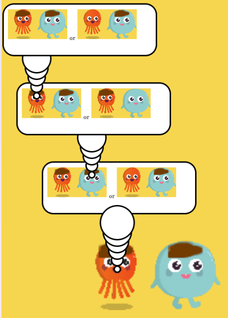

Minesweeper easy


Flatland

Philosophy

Outline

Open-source project



http://hintikkasworld.irisa.fr/

https://gitlab.inria.fr/fschwarz/hintikkasworld

[demo IJCAI-ECAI 2018]

- Web app
- Modular source code in Typescript
- Easy to add new examples
- Several contributors

Please contribute

- Coding
- Propose ideas and improvements

Outline

Outline

Epistemic states

Let $AP = \{p, p_1, \ldots\}$ be a countable set of atomic propositions.
Let $AGT = \{a, b, c, \ldots\}$ be a finite set of agents.

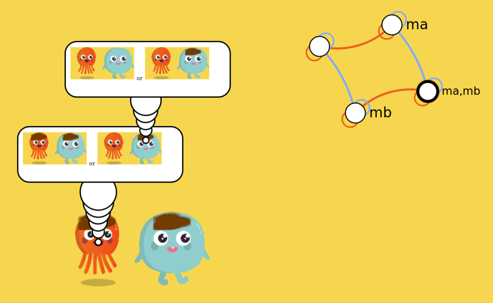
Definition

An epistemic model $\mathcal{M} = (W, (R_a)_{a \in AGT}, V)$ is a tuple where:

- $W = \{w, u, \ldots\}$ is a non-empty set of possible *worlds*;
- $R_a \subseteq W \times W$ is an *accessibility relation* for agent a;
- $V : W \to 2^{AP}$ is a *valuation function*.

A pair $(\mathcal{M}, w)$ is called a epistemic state, where w represents the actual world.

Example of an epistemic state



In Hintikka's World: Muddy children

- $W = \{w, u, v, s\}$;
- $R_a = \{(w, w), (w, u), (u, w), (u, u), (v, v), (v, s), (s, v), (s, s)\}$;
- $R_b = \{(w, w), (w, v), (v, w), (v, v), (u, u), (u, s), (s, u), (s, s)\}$;
- $V(w) = \{m_a, m_b\}$; $V(u) = \{m_b\}$; $V(v) = \{m_a\}$; $V(s) = \emptyset$.

Outline

Syntax of $\mathcal{L}_{\mathbf{EL}}$

Definition

The syntax of $\mathcal{L}_{\mathbf{EL}}$ is given by the following grammar:

$$\varphi, \psi, \ldots \quad ::= \quad p \mid \neg\varphi \mid (\varphi \vee \psi) \mid K_a\varphi$$

where p ranges over AP and a ranges over AGT.

The size of φ is the number of symbols needed to write φ.

Notation

$(\varphi \wedge \psi)$	*for* $\neg(\neg\varphi \vee \neg\psi)$;
$\hat{K}_a\varphi$	*for* $\neg K_a \neg\varphi$
$(\varphi \rightarrow \psi)$	*for* $(\neg\varphi \vee \psi)$

- $K_a\varphi$ is read 'agent a knows/believes that φ is true';
- $\hat{K}_a\varphi$ is read 'agent a considers φ as possible'.

Semantics of $\mathcal{L}_{\mathbf{EL}}$

Definition

The semantics of $\mathcal{L}_{\mathbf{EL}}$ is defined as follows:

$$
\begin{array}{ll}
\mathcal{M}, w \models p & \text{if } p \in V(w); \\
\mathcal{M}, w \models \neg\varphi & \text{if it is not the case that } \mathcal{M}, w \models \varphi; \\
\mathcal{M}, w \models (\varphi \vee \psi) & \text{if } \mathcal{M}, w \models \varphi \text{ or } \mathcal{M}, w \models \psi; \\
\mathcal{M}, w \models K_a\varphi & \text{if for all } u \text{ s.t. } wR_a u,\ \mathcal{M}, u \models \varphi
\end{array}
$$

Dual operators

$$\mathcal{M}, w \models K_a\varphi \quad \text{if for all } u \text{ s.t. } wR_au,\ \mathcal{M}, u \models \varphi$$

$$\mathcal{M}, w \models \hat{K}_a\varphi \quad \text{if there exists } u \text{ s.t. } wR_au \text{ and } \mathcal{M}, u \models \varphi.$$

$\mathcal{M}, w \models K_a m_b$

$\mathcal{M}, w \models \hat{K}_a m_a$

Practical session

In Hintikka's World: check formulas on the example you like

Syntax of formulas in Hintikka's world

`p`	
`(not phi)`	
`(phi or psi)`	
`(phi or phi or chi or ...)`	
`(phi and psi and chi or...)`	
`(K a phi)`	agent *a* knows/believes φ
`(Kpos a phi)`	agent *a* considers φ as possible

Example

`( (K a (p or q)) and (Kpos a r) )`

Common knowledge

Common knowledge of φ among agents in group G

Definition

The syntax of $\mathcal{L}_{\mathbf{ELCK}}$ is given by the following grammar:

$$\varphi \quad ::= \quad p \mid \neg\varphi \mid (\varphi \vee \varphi) \mid K_a\varphi \mid C_G\varphi$$

where p ranges over AP, a ranges over AGT, and G ranges over 2^{AGT}.

Definition

The semantics of $\mathcal{L}_{\mathbf{ELCK}}$ extended by the following clause:

- $\mathcal{M}, w \models C_G\varphi$ if for all $u \in W, wR_Gu$ implies $\mathcal{M}, u \models \varphi$ where R_G is the reflexive transitive closure of $\bigcup_{a \in G} R_a$.

Outline

Outline

Model checking problem

Definition (model checking problem)

- Input:
  - An epistemic state 
  - A formula, e.g. $K_a p$;
- Output: yes if  satisfies $K_a p$; no otherwise.

Model checking problem

Definition

The *model checking problem* is defined as follows.

- Input:
  - An epistemic state $\mathcal{M}, w$;
  - A formula φ;
- Output: yes if $\mathcal{M}, w \models \varphi$; no otherwise.

Theorem

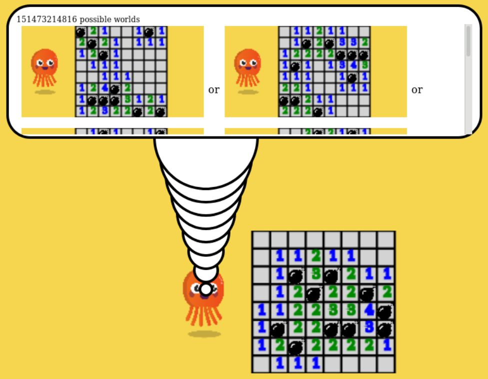
Model checking problem is P*-complete.*

Model checking algorithm

input: a Kripke model $\mathcal{M}$, a formula φ
output: the set of worlds in $\mathcal{M}$ in which φ holds
function mc($\mathcal{M}, \varphi$)
  match φ **do**
    case p :
      return $\{w \mid p$ holds in $\mathcal{M}, w\}$
    case $\neg\psi$:
      return $\overline{\mathrm{mc}(\mathcal{M}, \psi)}$
    case $(\psi_1 \vee \psi_2)$:
      return $\mathrm{mc}(\mathcal{M}, \psi_1) \cup \mathrm{mc}(\mathcal{M}, \psi_2)$
    case $K_a\psi$:
      return $\{w \mid R_a(w) \subseteq \mathrm{mc}(\mathcal{M}, \psi)\}$

Outline

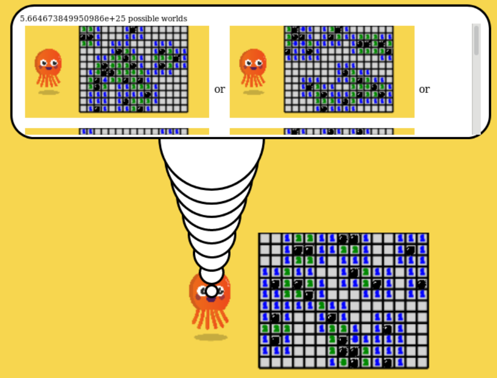
State explosion problem



Example

Minesweeper easy 8×8 with 10 bombs: $> 10^{12}$ possible worlds.

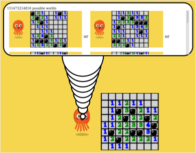
State explosion problem



Example

Minesweeper 10×12 with 20 bombs: $> 10^{25}$ possible worlds.

Solution to the state explosion problem



[**van Benthem**; **et al. 2015**], [**van Benthem et al. 2018**]

[Charrier _ AAMAS 2017], [Charrier _ AiML 2018]

- Succinct representations of epistemic states; **and** actions;
- Easy to specify by means of accessibility programs;
- Succinct model checking PSPACE-complete.

Outline

Outline

Satisfiability and validity

Definition

- A formula φ is *satisfiable* if there is an epistemic state $\mathcal{M}, w$ such that $\mathcal{M}, w \models \varphi$.
- A formula φ is *valid/a theorem* if for all epistemic states $\mathcal{M}, w$, we have $\mathcal{M}, w \models \varphi$.

Example

- $K_a p$ is satisfiable, but not valid.
- $(K_a p \quad \wedge \quad K_a(p \rightarrow q)) \quad \rightarrow \quad K_a q$ is valid.

Dual properties

φ is a theorem iff $\neg\varphi$ is not satisfiable.

Outline

Axiomatization

	all classical tautologies
Axiom K:	$K_a(\varphi \to \psi) \to (K_a\varphi \to K_a\psi)$
Modus ponens rule:	From φ and $\varphi \to \psi$, infer ψ
Necessitation rule:	From φ infer $K_a\varphi$

Theorem

A formula is a theorem iff it is provable in the axiomatization above.

[Blackburn et al. Modal logic, 2001]

Example

$K_a(\varphi \wedge \psi) \to K_a\varphi$ is theorem:

1. $(\varphi \wedge \psi) \to \varphi$ — classical tautology
2. $K_a((\varphi \wedge \psi) \to \varphi)$ — by necessitation rule on 1
3. $K_a((\varphi \wedge \psi) \to \varphi) \to (K_a(\varphi \wedge \psi) \to K_a\varphi)$ — Axiom K
4. $K_a(\varphi \wedge \psi) \to K_a\varphi$ — by modus ponens on 2, 3

Motivation of axiomatization

- the computation of knowledge is modeled;
- enables to explain why an agent knows sth;
  (link with justification logic)
- axiomatization helps to understand the principle of the logics
- we do not have to design a specific epistemic state, as in model checking

Outline

Classes of epistemic states

	Properties		Related axioms
K	all		
T	reflexive		$K_a\varphi \to \varphi$
D	seriality		$\hat{K}_a\top$
4	transitivity		$K_a\varphi \to K_aK_a\varphi$
5	Euclideanity		$\neg K_a\varphi \to K_a\neg K_a\varphi$

In Hintikka's World: Classes of models

Definition

A formula φ is a KD45-theorem if for all epistemic states $\mathcal{M}, w$ in which relations are serial, transitive and Euclidean, we have $\mathcal{M}, w \models \varphi$.

Theorem

A formula φ is a KD45-theorem iff it is provable in the axiomatisation above plus axioms D, 4, 5. [Sahlqvist, 1975]

Important classes: KD45 and S5 = KT45

Example (KD45, i.e. beliefs)

A formula φ is a KD45-theorem if for all epistemic states $\mathcal{M}, w$ in which relations are serial, transitive and Euclidean, we have $\mathcal{M}, w \models \varphi$.

Example (S5 = KT45, i.e. knowledge)

A formula φ is a S5-theorem if for all epistemic states $\mathcal{M}, w$ in which relations are equivalence relations, we have $\mathcal{M}, w \models \varphi$.

Outline

Complexity of theorem proving

Theorem

Without common knowledge:

	one single agent	*several agents*
K	PSPACE-*complete*	PSPACE-*complete*
KD45, S5	NP-*complete*	PSPACE-*complete*

With common knowledge (several agents): EXPTIME-*complete.*

[Halpern, Moses, A guide to completeness and complexity for modal logics of knowledge and belief. 1996]

Model checking more practical than theorem proving [Halpern, Vardi, 1991]

Outline

Outline

Strictly more expressive

Definition

Two formulas φ and ψ are *equivalent* if for all pointed models $\mathcal{M}, w$,

$$(\mathcal{M}, w \models \varphi) \text{ iff } (\mathcal{M}, w \models \psi)$$

Theorem

$\mathcal{L}_{\mathbf{ELCK}}$ is strictly more expressive than $\mathcal{L}_{\mathbf{EL}}$: no formula in $\mathcal{L}_{\mathbf{EL}}$ is equivalent to $C_{\{a,b\}}p$.

- By contradiction, suppose that φ in $\mathcal{L}_{\mathbf{EL}}$ is equivalent to $C_{\{a,b\}}p$;
- Let d be the modal depth of φ, e.g. $d = 3$;
- Let us consider the two models of
  In Hintikka's World: Language with Common knowledge is more expressive
- φ has the same value in both while $C_{\{a,b\}}p$ not.

Equally expressive

We may add in the language operators $E_G\varphi$ read as 'agents in G know φ':

- $\mathcal{M}, w \models E_G\varphi$ if for all agents $a \in G$, $\mathcal{M}, w \models K_a\varphi$.

Theorem

The language $\mathcal{L}_{\mathbf{EL}}$ augmented with the E_G's is equally expressive than $\mathcal{L}_{\mathbf{EL}}$:

$$E_G\varphi \equiv \bigwedge_{a \in G} K_a\varphi$$

Outline

Succinctness

Theorem

The language $\mathcal{L}_{\mathbf{EL}}$ augmented with the E_G's is exponentially more succinct than $\mathcal{L}_{\mathbf{EL}}$.

- $E_{\{a,b\}}E_{\{a,b\}}E_{\{a,b\}}\varphi \equiv K_aK_aK_a\varphi \wedge K_aK_aK_b\varphi \wedge K_aK_bK_a\varphi \wedge K_aK_bK_b\varphi \wedge K_bK_aK_a\varphi \wedge K_bK_aK_b\varphi \wedge K_bK_bK_a\varphi \wedge K_bK_bK_b\varphi$
- $E_{\{a,b\}} \dots E_{\{a,b\}}\varphi \equiv \dots$

Proof is involved: see [French, van der Hoek, Illiev, Kooi, AIJ 2013]