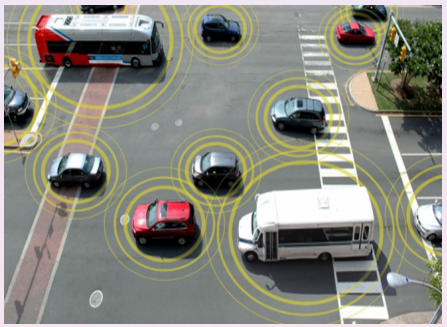
Intelligent artificial agents that detect and produce lies

François Schwarzentruber

Lorentz Center

13th January 2017

Potential applications



Outline

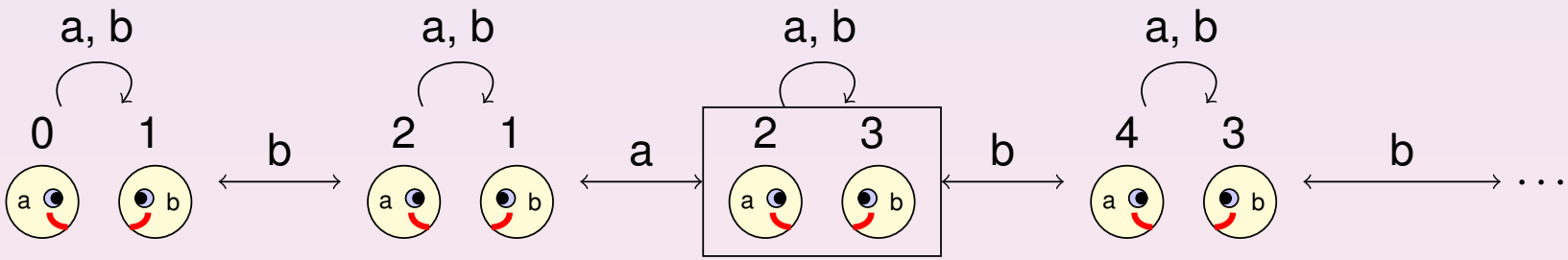
Outline

Prototype

http://people.irisa.fr/Francois.Schwarzentruber/hintikkasworld/

Outline

Initial situation described with a pointed Kripke model

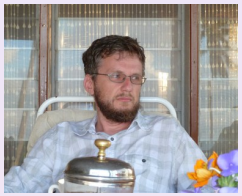


Outline

Actions described by event models


Alexandru Baltag


Lawrence S. Moss


Slawomir Solecki

Example (Public announcement "*a*'s number is 3")

Outline

Properties expressed in epistemic logic

Language constructions

a's number is 3 | *a*'s number is 4 | *b*'s number is 7 | . . .

not ...

(... **or** ...)

(... **and** ...)

(... → ...)

(... **knows** ...)

Example

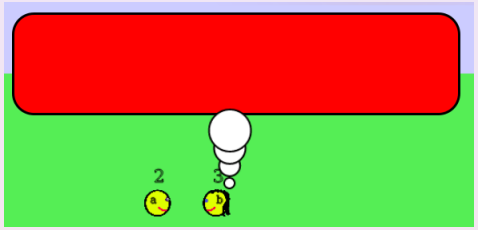
((*a* **knows** *b*'s number is 3) **and not** (*b* **knows** *a*'s number is 4))

Outline

Outline

Detecting some lies

Let us assume that agents always consider the real world as possible if no lies.



if *a*'s beliefs are inconsistent, then agent *a* has detected a lie.

model checking

Outline

Model checking definition

- Input: descriptions of an initial situation, a sequence of actions $\alpha_1, \ldots, \alpha_k$, a property;
- Does *property* holds after executing sequence of actions $\alpha_1, \ldots, \alpha_k$ from the initial situation?

initial situation α_1 α_2 $\ldots$ α_k property ✓

Outline

Outline

Epistemic planning

- Input: descriptions of an initial situation, a set of actions $\{\alpha_1, \ldots, \alpha_k\}$, a goal property;
- Does there exists a plan of actions $\alpha_{i_1}, \ldots, \alpha_{i_n}$ such that the goal property holds after executing the plan from initial situation?

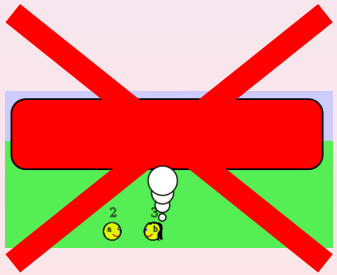
initial situation α_{i_1} α_{i_2} $\ldots$ α_{i_n} goal property ✓

Outline

Producing lies

It reduces to epistemic planning:

- set of actions $\{\alpha_1, \ldots, \alpha_k\}$ are the repertoire of the liar (it may contain lie actions and other actions);
- The goal property is the goal of the liar plus consistent of beliefs so that he will not been detected.



initial situation α_{i_1} α_{i_2} $\ldots$ α_{i_n} property ✓

Outline

Future work

- Add examples:
  - Modelling first and second-order examples of Torben Braüner;
  - Illustrating true lies shown by Thomas Agotnes;
- Implementation:
  - connecting with the model checker DEMO developped by Jan van Eijck and Malvin Gattinger
  - decidable fragments for planning
- Modeling: intentions (see Chiaki Sakama et al.), strategic reasoning, lyers *à la Yanjing* etc.
- Applications: present and discuss with psychiatrists for making a software for kids.

Every day research

Type of actions	Complexity of the model checking problem	Succinct version
Public announcements	P-complete	PSPACE-complete in EXPSPACE for a geometric version
General DEL-actions	PSPACE-complete	PSPACE-complete [new]

Type of actions	Complexity of the planning problem
Public announcements	NP-complete
Purely epistemic and propositional messages	PSPACE-complete
Propositional messages and effects	in k-EXPTIME
General DEL-actions	undecidable

k = modal depth of the goal property

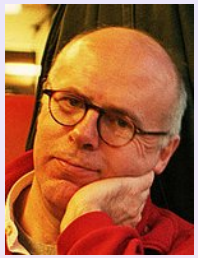
Special thanks


Guillaume Aucher


Andreas Herzig


Hans van Ditmarsch


Yanjing Wang

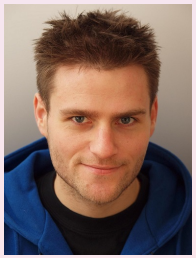

Jan van Eijck


Malvin Gattinger

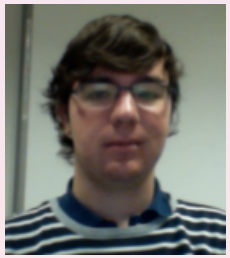

Mikkel Birkegaard Andersen


Thomas Bolander


Bastien Maubert


Martin Holm Jensen


Sophie Pinchinat

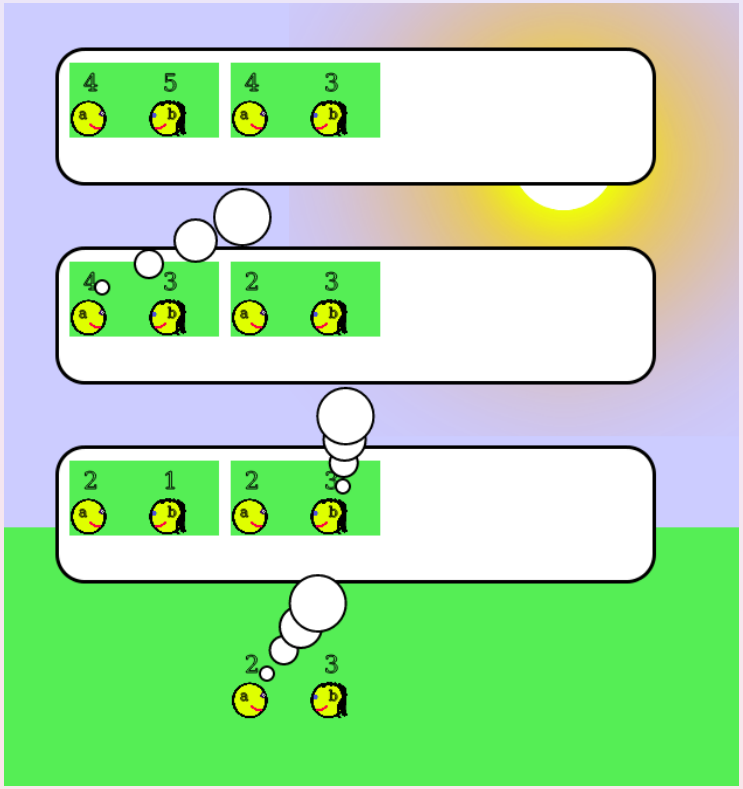

Tristan Charrier


Quan Yu


Ximing Wen


Yongmei Liu

Trugarez bras. Merci. Thank you. Dank u wel.



Feel free to use it!

`http://people.irisa.fr/Francois.Schwarzentruber/hintikkasworld/`