Intelligent artificial agents that detect and produce lies

François Schwarzentruber

Lorentz Center

13th January 2017
Potential applications
Outline

1. Intelligent agents with higher-order knowledge
   - Prototype
   - Initial situation
   - Actions
   - Properties

2. Model checking and detecting some lies

3. Epistemic planning and producing lies

4. Conclusion
Outline

1. Intelligent agents with higher-order knowledge
   - Prototype
     - Initial situation
     - Actions
     - Properties

2. Model checking and detecting some lies

3. Epistemic planning and producing lies

4. Conclusion
Prototype

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

1. Intelligent agents with higher-order knowledge
   - Prototype
   - Initial situation
   - Actions
   - Properties

2. Model checking and detecting some lies

3. Epistemic planning and producing lies

4. Conclusion
Initial situation described with a pointed Kripke model
Outline

1. Intelligent agents with higher-order knowledge
   - Prototype
   - Initial situation
   - Actions
   - Properties

2. Model checking and detecting some lies

3. Epistemic planning and producing lies

4. Conclusion
Actions described by event models

Example (Public announcement “a’s number is 3”)

true -> a,b -> a’s number is 3
Outline

1. Intelligent agents with higher-order knowledge
   - Prototype
   - Initial situation
   - Actions
   - Properties

2. Model checking and detecting some lies

3. Epistemic planning and producing lies

4. Conclusion
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 ...)
- (... → ...)

Example

```
(((a knows b’s number is 3) and not (b knows a’s number is 4)))
```
Outline

1. Intelligent agents with higher-order knowledge
2. Model checking and detecting some lies
   - Detecting some lies
   - Model checking definition
3. Epistemic planning and producing lies
4. Conclusion
Outline

1. Intelligent agents with higher-order knowledge
2. Model checking and detecting some lies
   - Detecting some lies
   - Model checking definition
3. Epistemic planning and producing lies
4. Conclusion
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.
Outline

1. Intelligent agents with higher-order knowledge
2. Model checking and detecting some lies
   - Detecting some lies
   - Model checking definition
3. Epistemic planning and producing lies
4. Conclusion
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 $\alpha_1 \; \alpha_2 \; \ldots \; \alpha_k$ property ✓
Outline

1. Intelligent agents with higher-order knowledge
2. Model checking and detecting some lies
3. Epistemic planning and producing lies
   - Epistemic planning
   - Producing lies
4. Conclusion
Outline

1. Intelligent agents with higher-order knowledge
2. Model checking and detecting some lies
3. Epistemic planning and producing lies
   - Epistemic planning
   - Producing lies
4. Conclusion
Epistemic planning

- Input: descriptions of an initial situation, a set of actions \( \{\alpha_1, \ldots, \alpha_k\} \), a goal property;
- Does there exist 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 \( \alpha_{i_1} \alpha_{i_2} \ldots \alpha_{i_n} \) goal property \( \checkmark \)
Outline

1. Intelligent agents with higher-order knowledge
2. Model checking and detecting some lies
3. Epistemic planning and producing lies
   - Epistemic planning
   - Producing lies
4. Conclusion
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** $\alpha_{i_1}, \alpha_{i_2}, \ldots, \alpha_{i_n}$ property $\checkmark$
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, liers à la Yanjing etc.
- Applications: present and discuss with psychiatrists for making a software for kids.
Every day research

<table>
<thead>
<tr>
<th>Type of actions</th>
<th>Complexity of the planning problem</th>
<th>Succinct version</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public announcements</td>
<td>NP-complete</td>
<td>PSPACE-complete [new]</td>
</tr>
<tr>
<td>Purely epistemic and propositional messages</td>
<td>PSPACE-complete</td>
<td>in EXPSPACE for a geometric version</td>
</tr>
<tr>
<td>Propositional messages and effects</td>
<td>in ( k )-EXPTIME</td>
<td></td>
</tr>
<tr>
<td>General DEL-actions</td>
<td>undecidable</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>( k = ) modal depth of the goal property</td>
<td></td>
</tr>
</tbody>
</table>
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

Feel free to use it!

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