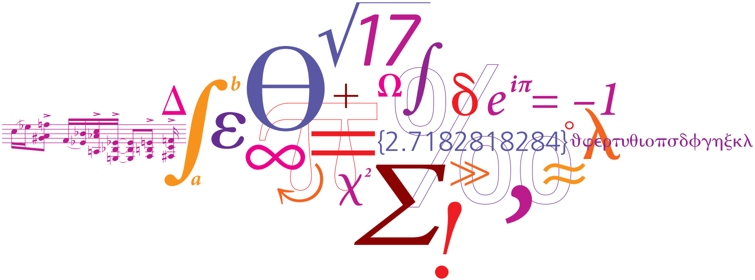
Implementing Theory of Mind on a Robot Using Dynamic Epistemic Logic

Thomas Bolander and Lasse Dissing, DTU Compute

LIRa talk, 25 June 2020



DTU Compute
Department of Applied Mathematics and Computer Science

Structure of the talk

- **Why** are we interested in Theory of Mind reasoning (social perspective-taking)?
- **What** did we do with it? Robot demo.
- **How** did we do it? Logic!

It is **applied logic**—applied to cognitive robotics and human-robot interaction. No hard theorems, no long proofs (sorry!).

Social AI: Why?

- Flexible and natural interaction with humans.
- Explainability: AI systems that can make themselves understood by humans (building trust).

Example of the necessity of social intelligence. Hospital robots in environments also inhabited by humans.

- *"I'm on the phone! If you say 'TUG has arrived' one more time I'm going to kick you in your camera."*

(Colin Barras, New Scientist, vol. 2738, 2009)

TUG hospital robot

The problem is general, so the solution has to be as well!

The 3 hardest problems in AI

Social intelligence: The ability to understand others and the social context effectively and thus to interact with other agents successfully.

Carl Frey, 2017
Kolding, Denmark

Toby Walsh, 2017
Science & Cocktails, Copenhagen

Both have **social intelligence** among the 3 human cognitive abilities that are hardest to simulate by computers and robots.

Social intelligence at work

A psychological experiment with an 18 months old kid. He didn't receive any instructions. (Warneken & Tomasello, 2006)

http://www2.compute.dtu.dk/~tobo/children_cabinet_trimmed.mov

Social intelligens: What is it?

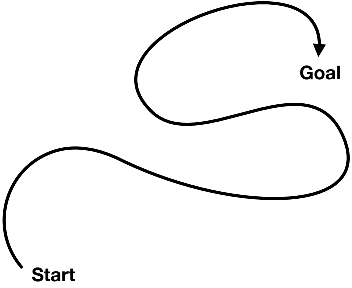
The child appears to have the ability to **put himself in the shoes of the adult**, understanding what he wants to achieve and what his abilities are.

Theory of Mind (ToM): The ability to understand and reason about the mental state of other agents, e.g. their beliefs, intentions and desires. (Premack & Woodruff, 1978)

Theory of Mind is essential to human social intelligence. (Baron-Cohen, 1997)

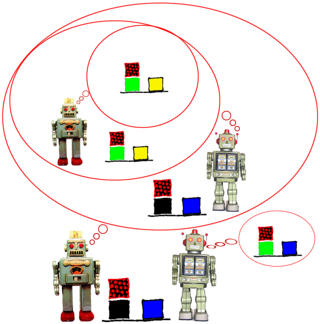
What would it take to make a robot do the same?

1. Ability to plan: **automated planning**. Check!
2. Ability to infer the goal of the adult: **goal recognition**. Only partly solved with current AI techniques.
3. Ability to take the perspective of the adult in the planning process: **epistemic planning**. Check! Epistemic planning = automated planning + dynamic epistemic logic [Bolander and Andersen, 2011].



Automated planning

$+$

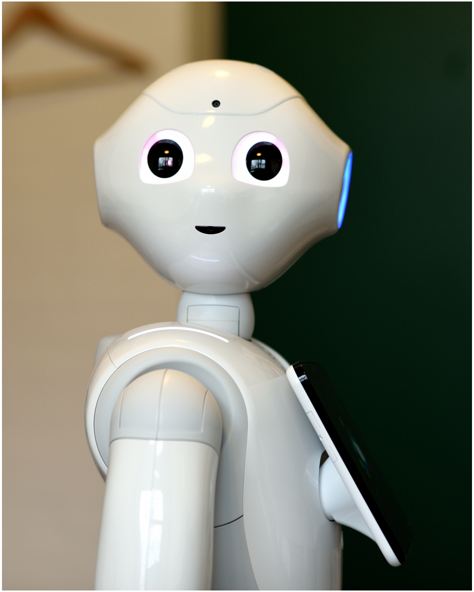


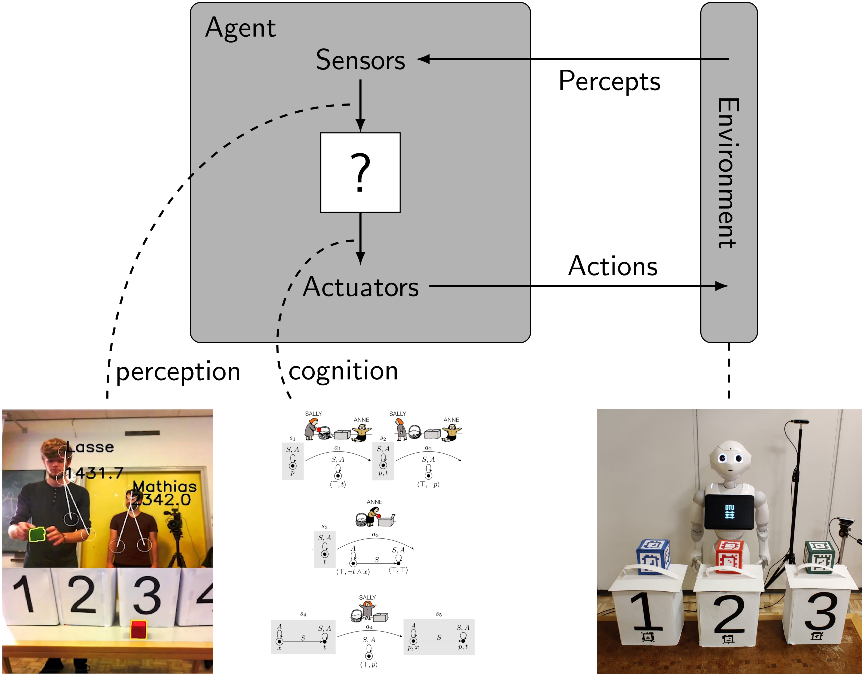
Logical reasoning about the mental states of other agents

A false-belief task: the Sally-Anne test

`http://www2.compute.dtu.dk/~tobo/sally_anne_trimmed.mp4`

Robot demo




Agent
Sensors
Percepts
Environment
?
Actions
Actuators
perception
cognition
Lasse
1431.7
Mathias
342.0
1
2
3
SALLY
ANNE
DTU
1
2
3

Wrongfully Accused by an Algorithm

In what may be the first known case of its kind, a faulty facial recognition match led to a Michigan man's arrest for a crime he did not commit.

The New York Times June 24, 2020

"This is not me," Robert Julian-Borchak Williams told investigators. "You think all Black men look alike?" Sylvia Jarrus for The New York Times

Perception via deep neural networks can never be 100% precise.

However, we can prove that the DEL formalism can correctly handle any expressible false-belief task of arbitrary order [Bolander, 2018]. So failure to pass a task is reduced to perception failures.

Perception layer: Detectors, world model and events

Detectors: Detect a specific kind of feature such as faces (dlib CNN face recognition), markers (AprilTag fiducial markers), and body poses (OpenPose).

Spatial world model: Keeps track of the spatial position of physical entities using the detectors. Physical entities are split into *objects* $\mathcal{O}$ and *agents* $\mathcal{A}$.

Events: The spatial world model informs other components in the system using *events*:

- $\texttt{Appear}(c)$: World model tracking locks to a new entity c.
- $\texttt{Disappear}(c)$: World model is no longer able to track c.
- $\texttt{pickup}(i, c)$: Agent i picks up object c. Triggered by hand of i entering bounding box of c.
- $\texttt{put}(i, c, b)$: Agent i puts object c in container b.

Cognition layer: Epistemic formulas and states

We use dynamic epistemic logic (DEL) with *postconditions*, *edge-conditioned action models* and *observability propositions* [Bolander, 2018].

Definition Let $\mathcal{O}$ and $\mathcal{A}$ be as above, and let Ψ be a set of predicates of first-order logic. The *epistemic language* $\mathcal{L}(\Psi, \mathcal{O}, \mathcal{A})$ is:

$$\phi ::= P(\omega) \mid i \lhd j \mid \neg\phi \mid \phi \wedge \phi \mid B_i \phi$$

where $i, j \in \mathcal{A}$, $P \in \Psi$ is a predicate of arity $ar(P) \in \mathbb{N}$, and $\omega \in (\mathcal{O} \cup \mathcal{A})^{ar(P)}$. Formulas $P(\omega)$ and $i \lhd j$ are *atoms*, and the set of these is denoted *Atm*.

Example. $B_{Anne} B_{Sally} In(marble, basket)$.

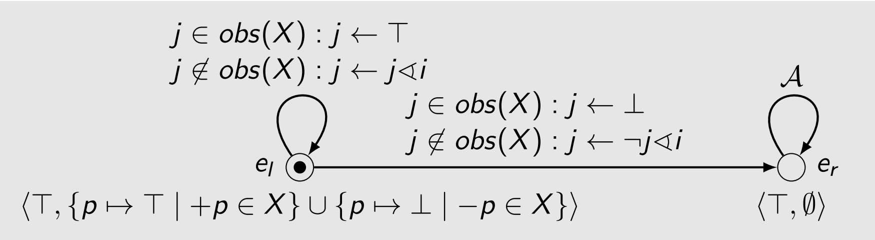
Semantics via epistemic states (Kripke models) as usual. No frame conditions (logic K).

Cognitive layer: Epistemic actions

Definition. An *action* is an expression $i{:}X$, where $i \in \mathcal{A}$ and X is a list (set) of assignments of the form $+p$ or $-p$ where $p \in Atm$.

Examples. $Anne{:}\ -In(marble, basket), +In(marble, box)$ (Anne moves marble from basket to box). $Sally{:}\ -Sally \lessdot Anne, -Anne \lessdot Sally$ (Sally leaves).

Action model for action $i{:}X$,
where $obs(X) = \{i \in \mathcal{A} \mid +i \lessdot j \in X \text{ or } -i \lessdot j \in X \text{ for some } j\}$



Note: All actions represented by the same generic edge-conditioned event model (simplified compared to [Bolander, 2018]).

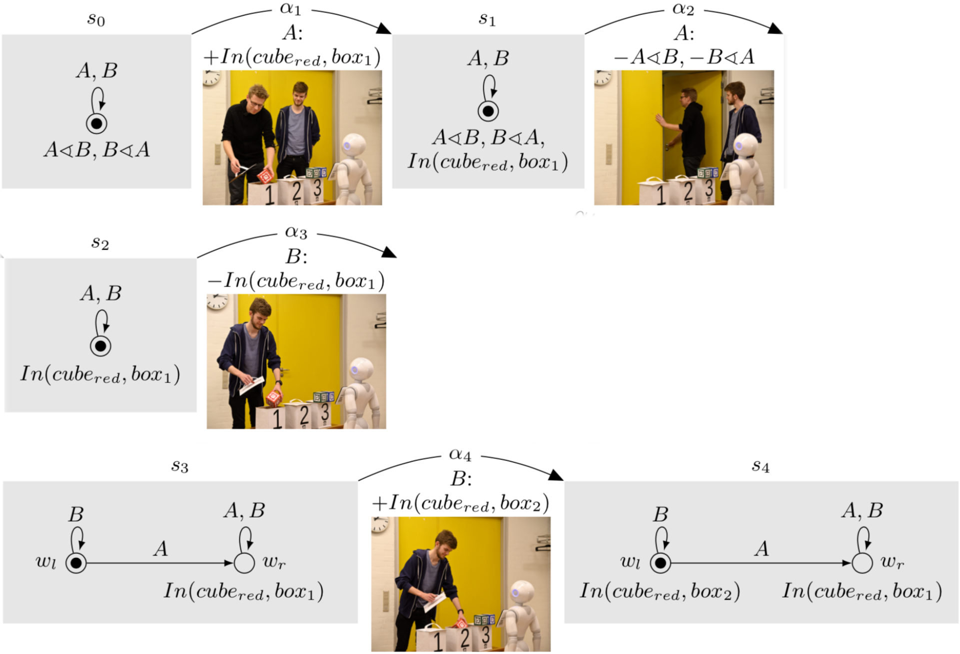
Sally-Anne on the robot

Performing the Sally-Anne story in front of the robot makes the perception system generate a sequence of events.

Each event is translated into an action (action model) applied to the previous epistemic state using product update. We maintain a set Φ of agents currently tracked by the robot.

- `pickup`(i, c, b). Apply action $i{:}{-}In(c, b)$.
- `put`(i, c, b). Apply action $i{:}{+}In(c, b)$.
- `Appear`(c)/`Disappear`(c). Update Φ, then apply action $i{:}\{+i \lessdot j \mid i, j \in \Phi\} \cup \{-j \lessdot k \mid (j, k) \in (\Phi \times (\mathcal{A} - \Phi)) \cup ((\mathcal{A} - \Phi) \times \Phi)\}$.

Note: Simplified treatment of observability change (only considering co-presence and absence).


s_0
A, B
$A \lhd B, B \lhd A$
α_1
A:
$+In(cube_{red}, box_1)$
s_1
A, B
$A \lhd B, B \lhd A,$
$In(cube_{red}, box_1)$
α_2
A:
$-A \lhd B, -B \lhd A$
s_2
A, B
$In(cube_{red}, box_1)$
α_3
B:
$-In(cube_{red}, box_1)$
s_3
B
A, B
A
w_l
w_r
$In(cube_{red}, box_1)$
α_4
B:
$+In(cube_{red}, box_2)$
s_4
B
A, B
A
w_l
w_r
$In(cube_{red}, box_2)$
$In(cube_{red}, box_1)$

Model queries

Definition. A *query* is a formula of $\mathcal{L}(\Psi, \mathcal{O}, \mathcal{A})$ where one or more constant symbols have been replaced by variables. We use standard notation $\phi(x_1, \ldots, x_n)$ for such formulas, where $\phi(c_1, \ldots, c_n)$ is the result of substituting c_i for x_i everywhere. The *answer* to a query $\phi(x_1, \ldots, x_n)$ in an epistemic state s is the formula
$\phi(x_1, \ldots, x_n)^s := \{(c_1, \ldots, c_n) \in (\mathcal{O} \cup \mathcal{A})^n \mid s \models \phi(c_1, \ldots, c_n)\}$.

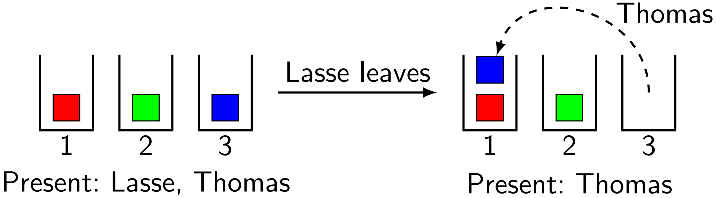
Speech input is first transcribed using DanSpeech (Danish) or Google Speech (English). The textual output is then parsed as a context-free language and transformed into an answer using a model query.

Example. The robot is in state s and asked "Where does Lasse believe that the red cube is?". The robot answers "Lasse believes the red cube is in $(B_{Lasse} In(cube_{red}, x))^s$".

Goal recognition and planning

- We add announcements, so the robot can be helpful by announcing facts.
- The robot does epistemic planning with implicit coordination: multi-agent planning with perspective shifts [Nebel et al., 2019, Bolander et al., 2018, Engesser et al., 2017].

Example. Consider the following action sequence:



If I say "I want two cubes in the same box" nothing happens. Lasse arrives and says the same. Now the robot replies: "It is already true".

Afterwards Lasse says: "I want three cubes in the same box". The robot replies: "Box 3 is empty".

Conclusion and future work

- We built a robotic system using deep learning, DEL and epistemic planning to pass false-belief tasks and make helpful announcements.
- We are the first to built a robotic system that can pass higher-order false-belief tasks. E.g. it managed to pass a second-order task that we didn't design it specifically for, and hadn't previously tested it on.

Future work. Most importantly, look at alternative DEL-like logical formalisms (mainly formalisms that can deal with belief revision):

- Plausibility models [Baltag and Smets, 2008]. Ideally with abduction (current work with Sonja Smets).
- Temporal visibility models [Solaki and Velázquez-Quesada, 2019].
- Extensions of the m$\mathcal{A}^*$ action language [Buckingham et al., 2020].
- DEL based on belief bases [Lorini, 2020].

References I

Baltag, A. and Smets, S. (2008).
A Qualitative Theory of Dynamic Interactive Belief Revision.
In Logic and the Foundations of Game and Decision Theory (LOFT7), (Bonanno, G., van der Hoek, W. and Wooldridge, M., eds), vol. 3, of Texts in Logic and Games pp. 13–60, Amsterdam University Press.

Bolander, T. (2018).
Seeing Is Believing: Formalising False-Belief Tasks in Dynamic Epistemic Logic.
In Jaakko Hintikka on Knowledge and Game-Theoretical Semantics pp. 207–236. Springer.

Bolander, T. and Andersen, M. B. (2011).
Epistemic Planning for Single- and Multi-Agent Systems.
Journal of Applied Non-Classical Logics *21*, 9–34.

Bolander, T., Engesser, T., Mattmüller, R. and Nebel, B. (2018).
Better Eager Than Lazy? How Agent Types Impact the Successfulness of Implicit Coordination.
In Proceedings of the 16th International Conference on Principles of Knowledge Representation and Reasoning (KR 2018) AAAI Press.

Buckingham, D., Kasenberg, D. and Scheutz, M. (2020).
Simultaneous Representation of Knowledge and Belief for Epistemic Planning with Belief Revision.
In International Conference on Principles of Knowledge Representation and Reasoning (KR 2020).

Engesser, T., Bolander, T., Mattmüller, R. and Nebel, B. (2017).
Cooperative Epistemic Multi-Agent Planning for Implicit Coordination.
In Proceedings of Methods for Modalities Electronic Proceedings in Theoretical Computer Science.

Lorini, E. (2020).
Rethinking epistemic logic with belief bases.
Artificial Intelligence *282*, 103233.

References II

Nebel, B., Bolander, T., Engesser, T. and Mattmüller, R. (2019).
Implicitly Coordinated Multi-Agent Path Finding under Destination Uncertainty: Success Guarantees and Computational Complexity.
Journal of Artificial Intelligence Research *64*, 497–527.

Solaki, A. and Velázquez-Quesada, F. R. (2019).
Towards a Logical Formalisation of Theory of Mind: A Study on False Belief Tasks.
In International Workshop on Logic, Rationality and Interaction pp. 297–312, Springer.