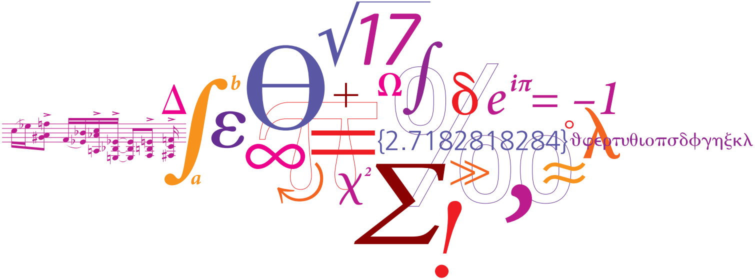
Seeing is Believing: Formalising False-Belief Tasks in Dynamic Epistemic Logic

Thomas Bolander, DTU Compute, Technical University of Denmark

Jaakko Hintikka Memorial Conference, Helsinki, 8 September 2016



DTU Compute
Department of Applied Mathematics and Computer Science

Social intelligence and anti-social robots

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

Social intelligence and anti-social robots

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

Frustrated users of hospital robots in USA:

- *"TUG was a hospital worker, and its colleagues expected it to have some social smarts, the absence of which led to frustration—for example, when it always spoke in the same way in both quiet and busy situations."*

TUG hospital robot

[Barras 2009]

Social intelligence and anti-social robots

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

Frustrated users of hospital robots in USA:

- *"TUG was a hospital worker, and its colleagues expected it to have some social smarts, the absence of which led to frustration—for example, when it always spoke in the same way in both quiet and busy situations."*
- *"I'm on the phone! If you say 'TUG has arrived' one more time I'm going to kick you in your camera."*

TUG hospital robot

[Barras 2009]

Social intelligence and anti-social robots

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

Frustrated users of hospital robots in USA:

- *"TUG was a hospital worker, and its colleagues expected it to have some social smarts, the absence of which led to frustration—for example, when it always spoke in the same way in both quiet and busy situations."*
- *"I'm on the phone! If you say 'TUG has arrived' one more time I'm going to kick you in your camera."*
- *"It doesn't have the manners we teach our children. I find it insulting that I stand out of the way for patients... but it just barrels right on."*

TUG hospital robot

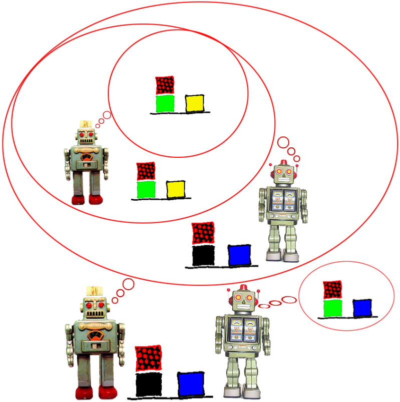
[Barras 2009]

Theory of Mind and false-belief tasks

Theory of Mind (ToM): The ability of attributing mental states—beliefs, intentions, desires, etc.—to other agents.

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

The strength of a human child's ToM is often tested with a **false-belief task** such as the **Sally-Anne task** [Wimmer and Perner, 1983].



Goal of the present work

Overall goal: To formalise false-belief tasks in a suitable logic.

Criteria for the formalisations:

- **Robustness**. *The formalism should not only be able to deal with one or two selected false-belief tasks, but with as many as possible, with no strict limit on the order of belief attribution.*
- **Faithfulness**. *Each action of the false-belief story should correspond to an action in the formalism in a natural way, and it should be fairly straightforward, not requiring ingenuity, to find out what that action of the formalism is. The formalisation of the false-belief story should only consist of these formalised actions.*

The ultimate aim:

- To provide the basis for a reasoning engine for artificial agents with ToM capabilities.

Comparison of false-belief task agents

The **Sally-Anne task** requires first-order belief attribution (attributing beliefs to Sally). Some false-belief tasks require ***n*-th order belief attribution** for $n > 1$.

Existing full formalisations/implementations of false-belief tasks:

	platform	h-o reas.	other features
CRIBB [Wahl and Spada, 2000]	Prolog	≤ 2	goal recognition, plan recognition
Edd Hifeng [Arkoudas and Bringsjord, 2008]	event calc.	≤ 1	Second Life avatar
Leonardo [Breazeal et al., 2011]	C5 agent arch.	≤ 1	goal recognition, learning
[Sindlar, 2011]	ext. of PDL, impl. in 2APL	≤ 1	goal recognition
ACT-R agent [Arslan et al., 2013]	ACT-R cogn. architecture	∞	learning
Hybrid logic agent [Braüner, 2013]	hybrid logic	∞	temporal reasoning

Structure of the talk

- Formalisation of the Sally-Anne task in standard Dynamic Epistemic Logic (DEL). (DEL because: 1) it can deal with arbitrary levels of higher-order reasoning (beliefs about beliefs); 2) arbitrary actions can explicitly be modelled).

Structure of the talk

- Formalisation of the Sally-Anne task in standard Dynamic Epistemic Logic (DEL). (DEL because: 1) it can deal with arbitrary levels of higher-order reasoning (beliefs about beliefs); 2) arbitrary actions can explicitly be modelled).
- Two problems with the standard DEL formalisation.

Structure of the talk

- Formalisation of the Sally-Anne task in standard Dynamic Epistemic Logic (DEL). (DEL because: 1) it can deal with arbitrary levels of higher-order reasoning (beliefs about beliefs); 2) arbitrary actions can explicitly be modelled).
- Two problems with the standard DEL formalisation.
- Extending DEL to provide better formalisations of false-belief tasks: *observability propositions* and *edge-conditioned event models*.

Structure of the talk

- Formalisation of the Sally-Anne task in standard Dynamic Epistemic Logic (DEL). (DEL because: 1) it can deal with arbitrary levels of higher-order reasoning (beliefs about beliefs); 2) arbitrary actions can explicitly be modelled).
- Two problems with the standard DEL formalisation.
- Extending DEL to provide better formalisations of false-belief tasks: *observability propositions* and *edge-conditioned event models*.
- Formalisation of Sally-Anne and the second-order chocolate task in the extended formalism.

Structure of the talk

- Formalisation of the Sally-Anne task in standard Dynamic Epistemic Logic (DEL). (DEL because: 1) it can deal with arbitrary levels of higher-order reasoning (beliefs about beliefs); 2) arbitrary actions can explicitly be modelled).
- Two problems with the standard DEL formalisation.
- Extending DEL to provide better formalisations of false-belief tasks: *observability propositions* and *edge-conditioned event models*.
- Formalisation of Sally-Anne and the second-order chocolate task in the extended formalism.
- Robustness and faithfulness revisited.

Structure of the talk

- Formalisation of the Sally-Anne task in standard Dynamic Epistemic Logic (DEL). (DEL because: 1) it can deal with arbitrary levels of higher-order reasoning (beliefs about beliefs); 2) arbitrary actions can explicitly be modelled).
- Two problems with the standard DEL formalisation.
- Extending DEL to provide better formalisations of false-belief tasks: *observability propositions* and *edge-conditioned event models*.
- Formalisation of Sally-Anne and the second-order chocolate task in the extended formalism.
- Robustness and faithfulness revisited.

I assume familiarity with epistemic logic, but not necessarily with dynamic epistemic logic.

Constants of modelling language

In the following we will use the following agent symbols:

- S: Sally.
- A: Anne.

We will use the following propositional symbols:

- *large*: The cube is in the large container.
- *small*: The cube is in the small container.
- *sally*: Sally is present in the room with Anne.

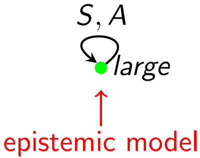
Dynamic Epistemic Logic (DEL) by example

We use the **event models** of DEL [Baltag et al., 1998] with added postconditions (ontic actions) as in [van Ditmarsch and Kooi, 2008].

Dynamic Epistemic Logic (DEL) by example

We use the **event models** of DEL [Baltag et al., 1998] with added postconditions (ontic actions) as in [van Ditmarsch and Kooi, 2008].

Example. Anne transfers the cube from the large to the small container in Sally's absence:

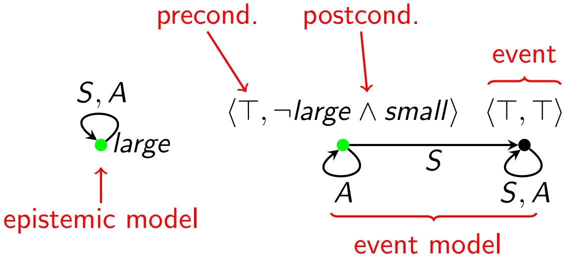


- **Epistemic models**: Multi-agent K models. We use green nodes (●) to denote the actual world.

Dynamic Epistemic Logic (DEL) by example

We use the **event models** of DEL [Baltag et al., 1998] with added postconditions (ontic actions) as in [van Ditmarsch and Kooi, 2008].

Example. Anne transfers the cube from the large to the small container in Sally's absence:

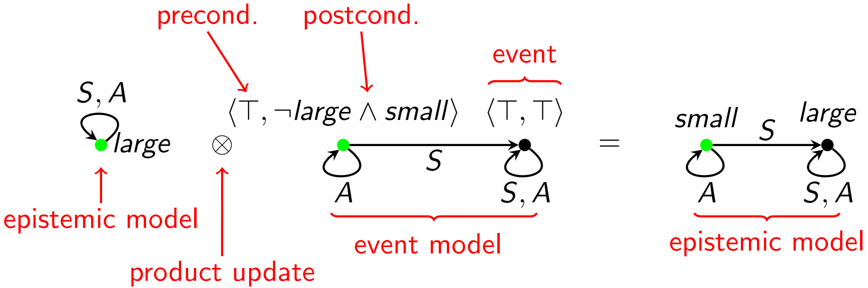


- **Epistemic models**: Multi-agent K models. We use green nodes (•) to denote the actual world.
- **Event model**: Represents the action of transferring the cube.

Dynamic Epistemic Logic (DEL) by example

We use the **event models** of DEL [Baltag et al., 1998] with added postconditions (ontic actions) as in [van Ditmarsch and Kooi, 2008].

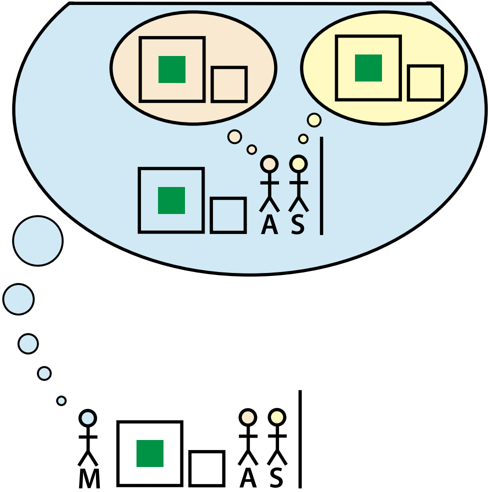
Example. Anne transfers the cube from the large to the small container in Sally's absence:



- **Epistemic models**: Multi-agent K models. We use green nodes (•) to denote the actual world.
- **Event model**: Represents the action of transferring the cube.
- **Product update**: The updated model represents the situation after the action has taken place.

Modelling Sally-Anne in DEL

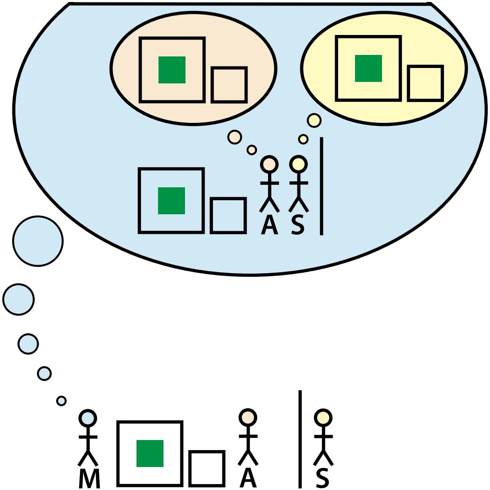
1. Sally has placed cube in large container:



$s_1 = \overset{S,\,A}{\circlearrowleft} \bullet \; large, sally$

Modelling Sally-Anne in DEL

2. Sally leaves room:

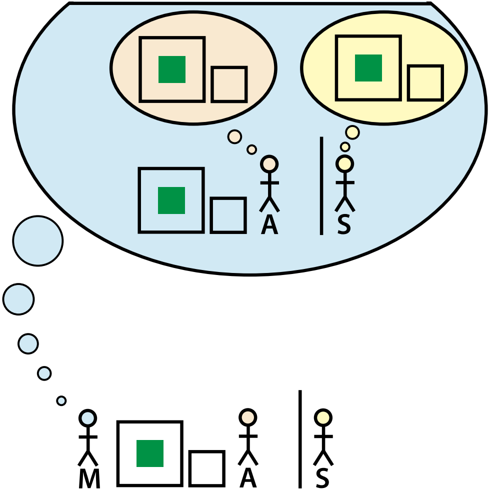


$$s_1 = \bullet^{S,A}\ \mathit{large}, \mathit{sally}$$

$$a_2 = \bullet^{S,A}\ \langle \top, \neg \mathit{sally} \rangle$$

Modelling Sally-Anne in DEL

2. Sally leaves room:



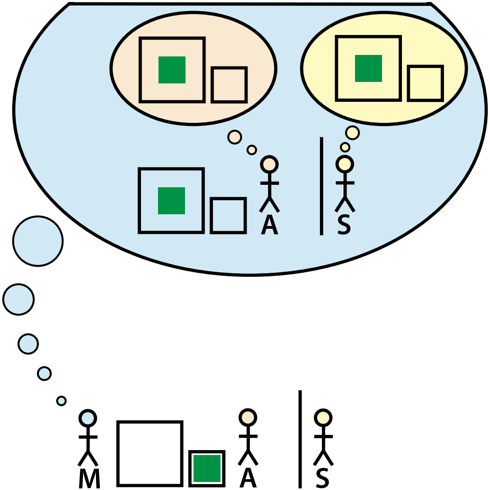
$s_1 = \circlearrowleft^{S,A} \bullet$ *large*, *sally*

$a_2 = \circlearrowleft^{S,A} \bullet \langle \top, \neg sally \rangle$

$s_2 = s_1 \otimes a_2 = \circlearrowleft^{S,A} \bullet$ *large*

Modelling Sally-Anne in DEL

3. Anne transfers cube to small container:



$$s_1 = \begin{tikzcd} \bullet\; large, sally \arrow[loop above, "S, A"] \end{tikzcd}$$

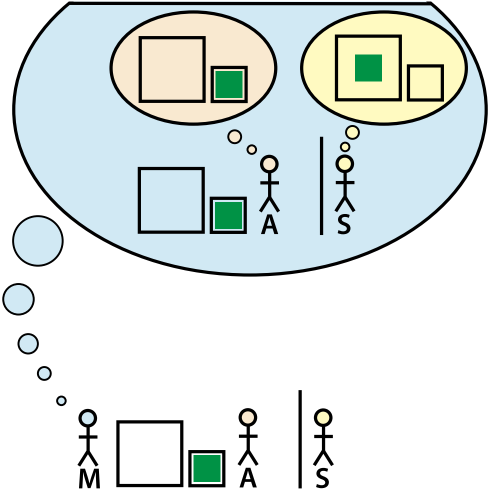
$$a_2 = \begin{tikzcd} \bullet\; \langle \top, \neg sally \rangle \arrow[loop above, "S, A"] \end{tikzcd}$$

$$s_2 = s_1 \otimes a_2 = \begin{tikzcd} \bullet\; large \arrow[loop above, "S, A"] \end{tikzcd}$$

$$a_3 = \begin{tikzcd} \underset{\langle \top, \neg large \wedge small \rangle}{\bullet} \arrow[loop above, "A"] \arrow[rr, "S"] & & \underset{\langle \top, \top \rangle}{\bullet} \arrow[loop above, "S, A"] \end{tikzcd}$$

Modelling Sally-Anne in DEL

3. Anne transfers cube to small container:



$$s_1 = \begin{tikzcd} \underset{\textit{large}, \textit{sally}}{\bullet} \arrow[loop above, "S, A"] \end{tikzcd}$$

$$a_2 = \begin{tikzcd} \underset{\langle \top, \neg \textit{sally} \rangle}{\bullet} \arrow[loop above, "S, A"] \end{tikzcd}$$

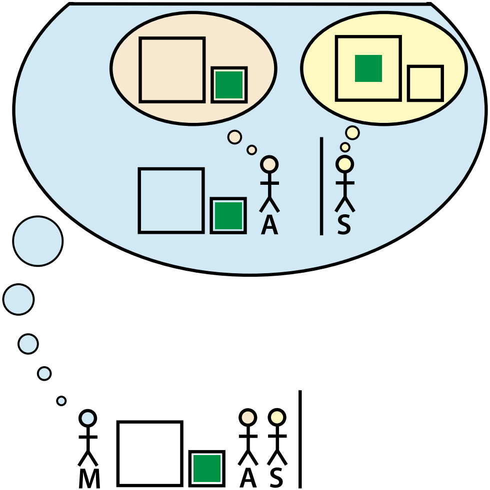
$$s_2 = s_1 \otimes a_2 = \begin{tikzcd} \underset{\textit{large}}{\bullet} \arrow[loop above, "S, A"] \end{tikzcd}$$

$$a_3 = \begin{tikzcd} \underset{\langle \top, \neg \textit{large} \wedge \textit{small} \rangle}{\bullet} \arrow[loop above, "A"] \arrow[rr, "S"] && \underset{\langle \top, \top \rangle}{\bullet} \arrow[loop above, "S, A"] \end{tikzcd}$$

$$s_3 = s_2 \otimes a_3 = \begin{tikzcd} \underset{\textit{small}}{\bullet} \arrow[loop above, "A"] \arrow[rr, "S"] && \underset{\textit{large}}{\bullet} \arrow[loop above, "S, A"] \end{tikzcd}$$

Modelling Sally-Anne in DEL

4. Sally re-enters:



$$s_1 = \overset{S,A}{\bullet}\ \mathit{large}, \mathit{sally}$$

$$a_2 = \overset{S,A}{\bullet}\ \langle \top, \neg \mathit{sally} \rangle$$

$$s_2 = s_1 \otimes a_2 = \overset{S,A}{\bullet}\ \mathit{large}$$

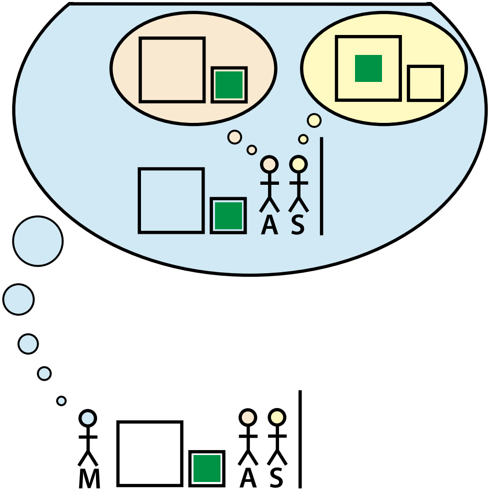
$$a_3 = \begin{tikzcd} \underset{\langle \top, \neg \mathit{large} \wedge \mathit{small} \rangle}{\bullet} \arrow[loop above, "A"] \arrow[r, "S"] & \underset{\langle \top, \top \rangle}{\bullet} \arrow[loop above, "{S,A}"] \end{tikzcd}$$

$$s_3 = s_2 \otimes a_3 = \begin{tikzcd} \underset{\mathit{small}}{\bullet} \arrow[loop above, "A"] \arrow[r, "S"] & \underset{\mathit{large}}{\bullet} \arrow[loop above, "{S,A}"] \end{tikzcd}$$

$$a_4 = \overset{S,A}{\bullet}\ \langle \top, \mathit{sally} \rangle$$

Modelling Sally-Anne in DEL

4. Sally re-enters:



$$s_1 = \begin{tikzcd} \underset{large,\, sally}{\bullet} \arrow[loop above, "{S,A}"] \end{tikzcd}$$

$$a_2 = \begin{tikzcd} \underset{\langle \top, \neg sally \rangle}{\bullet} \arrow[loop above, "{S,A}"] \end{tikzcd}$$

$$s_2 = s_1 \otimes a_2 = \begin{tikzcd} \underset{large}{\bullet} \arrow[loop above, "{S,A}"] \end{tikzcd}$$

$$a_3 = \begin{tikzcd} \underset{\langle \top, \neg large \wedge small \rangle}{\bullet} \arrow[loop above, "A"] \arrow[r, "S"] & \underset{\langle \top, \top \rangle}{\bullet} \arrow[loop above, "{S,A}"] \end{tikzcd}$$

$$s_3 = s_2 \otimes a_3 = \begin{tikzcd} \underset{small}{\bullet} \arrow[loop above, "A"] \arrow[r, "S"] & \underset{large}{\bullet} \arrow[loop above, "{S,A}"] \end{tikzcd}$$

$$a_4 = \begin{tikzcd} \underset{\langle \top, sally \rangle}{\bullet} \arrow[loop above, "{S,A}"] \end{tikzcd}$$

$$s_4 = s_3 \otimes a_4 =$$

$$\begin{tikzcd} \underset{small,\, sally}{\bullet} \arrow[loop above, "A"] \arrow[r, "S"] & \underset{large,\, sally}{\bullet} \arrow[loop above, "{S,A}"] \end{tikzcd}$$

Modelling Sally-Anne in DEL

1. Sally has placed cube in large container: $s_1 = \begin{tikzcd} \underset{large,\, sally}{\bullet} \arrow[loop above, "S,A"] \end{tikzcd}$
2. Sally leaves the room: $a_2 = \begin{tikzcd} \underset{\langle \top, \neg sally \rangle}{\bullet} \arrow[loop above, "S,A"] \end{tikzcd}$
3. Anne transfers cube: $a_3 = \begin{tikzcd} \underset{\langle \top, \neg large \wedge small \rangle}{\bullet} \arrow[loop above, "A"] \arrow[r, "S"] & \underset{\langle \top, \top \rangle}{\bullet} \arrow[loop above, "S,A"] \end{tikzcd}$
4. Sally re-enters: $a_4 = \begin{tikzcd} \underset{\langle \top, sally \rangle}{\bullet} \arrow[loop above, "S,A"] \end{tikzcd}$

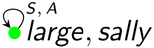
$$s_4 = s_1 \otimes a_2 \otimes a_3 \otimes a_4 = \begin{tikzcd} \underset{small,\, sally}{\bullet} \arrow[loop above, "A"] \arrow[r, "S"] & \underset{large,\, sally}{\bullet} \arrow[loop above, "S,A"] \end{tikzcd}$$

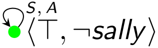
We have:

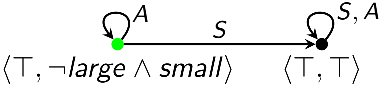
$$s_4 \models B_S\, large$$

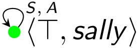
Thus the modeller will answer the question "where does Sally believe the cube is" with "in the large container", hence passing the Sally-Anne test!

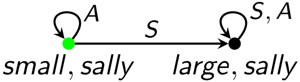
Modelling Sally-Anne in DEL

1. Sally has placed cube in large container: $s_1 =$ 

2. Sally leaves the room: $a_2 =$ 

3. Anne transfers cube: $a_3 =$ 

4. Sally re-enters: $a_4 =$ 

$s_4 = s_1 \otimes a_2 \otimes a_3 \otimes a_4 =$ 

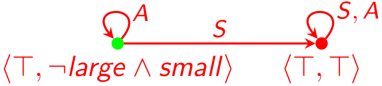
We have:

$$s_4 \models B_S\, large$$

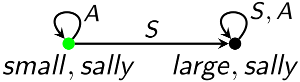
Thus the modeller will answer the question "where does Sally believe the cube is" with "in the large container", hence passing the Sally-Anne test!

Now note that $s_1 \otimes a_3 = s_4$. Thus Sally leaving and re-entering doesn't have any effect on the model the agent ends up with! Something is not right!...

Modelling Sally-Anne in DEL

1. Sally has placed cube in large container: $s_1 = \overset{S,A}{\bullet}\, \mathit{large}, \mathit{sally}$
2. Sally leaves the room: $a_2 = \overset{S,A}{\bullet}\, \langle \top, \neg \mathit{sally} \rangle$
3. Anne transfers cube: $a_3 =$ 

4. Sally re-enters: $a_4 = \overset{S,A}{\bullet}\, \langle \top, \mathit{sally} \rangle$

$s_4 = s_1 \otimes a_2 \otimes a_3 \otimes a_4 =$ 

We have:

$$s_4 \models B_S\, \mathit{large}$$

Thus the modeller will answer the question "where does Sally believe the cube is" with "in the large container", hence passing the Sally-Anne test!

Now note that $s_1 \otimes a_3 = s_4$. Thus Sally leaving and re-entering doesn't have any effect on the model the agent ends up with! Something is not right!...

Two problems

The current formalisation has two problems:

1. Even if Sally doesn't leave the room, she still gets the false belief.
2. The formalisation is not *faithful*: How did we get from the informal action descriptions to the event models?

Solving the two problems

To solve both problems of the previous slide, we add two new building blocks to DEL:

1. **Observability propositions**. A new set of propositional symbols of the form $i \lhd j$ (i sees j). $S \lhd A$: Sally is observing the actions of Anne. Inspired by [van Ditmarsch et al., 2013, Seligman et al., 2013].

Solving the two problems

To solve both problems of the previous slide, we add two new building blocks to DEL:

1. **Observability propositions**. A new set of propositional symbols of the form $i \triangleleft j$ (i sees j). $S \triangleleft A$: Sally is observing the actions of Anne. Inspired by [van Ditmarsch et al., 2013, Seligman et al., 2013].
2. **Edge-conditioned event models**. Edges of event models are generalised to $e_1 \xrightarrow{i:\phi} e_2$ meaning: agent i has an edge from e_1 to e_2 if ϕ is true (at e_1).

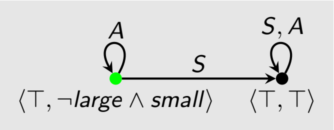
Solving the two problems

To solve both problems of the previous slide, we add two new building blocks to DEL:

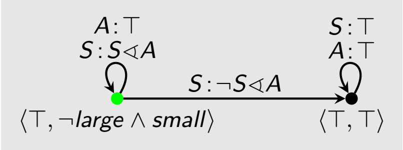
1. **Observability propositions**. A new set of propositional symbols of the form $i \lhd j$ (i sees j). $S \lhd A$: Sally is observing the actions of Anne. Inspired by [van Ditmarsch et al., 2013, Seligman et al., 2013].
2. **Edge-conditioned event models**. Edges of event models are generalised to $e_1 \xrightarrow{i\,:\,\phi} e_2$ meaning: agent i has an edge from e_1 to e_2 if ϕ is true (at e_1).

Putting the new building blocks together, the action of Anne transferring the cube becomes:

Before:



After:

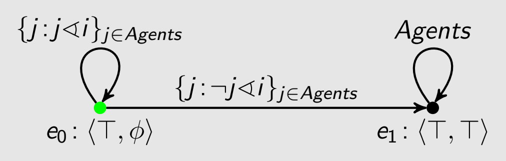


Generic edge-conditioned event models

We also get closer to *faithfulness*: "Who observes what" no longer has to be encoded explicitly in the structure of the event model, so all ontic actions can be represented by the same generic action type $do(i, \phi)$.

ontic action $do(i, \phi)$: agent i makes ϕ true (where ϕ is a conjunction of propositional literals). **Example**: $do(A, \neg large \wedge small)$.

event model for $do(i, \phi)$

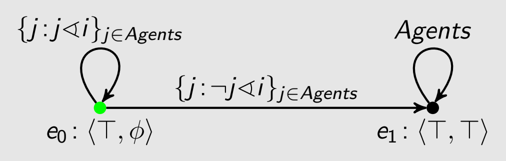


Generic edge-conditioned event models

We also get closer to *faithfulness*: “Who observes what” no longer has to be encoded explicitly in the structure of the event model, so all ontic actions can be represented by the same generic action type $do(i, \phi)$.

ontic action $do(i, \phi)$: agent i makes ϕ true (where ϕ is a conjunction of propositional literals). **Example**: $do(A, \neg large \wedge small)$.

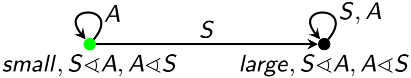
event model for $do(i, \phi)$



Observability changing action $oc(\phi)$: ϕ is made true, where ϕ is a conjunction of observation literals (observation propositions and their negation). **Example**: $oc(\neg S \triangleleft A \wedge \neg A \triangleleft S)$. (Event model omitted).

Modelling Sally-Anne in the new language

1. Sally has placed cube in large container: $s_1 = \overset{S,A}{\circlearrowleft}\, \mathit{large}, S \triangleleft A, A \triangleleft S$
2. Sally leaves the room: $a_2 = oc(\neg S \triangleleft A \wedge \neg A \triangleleft S)$
3. Anne transfers cube: $a_3 = do(A, \neg \mathit{large} \wedge \mathit{small})$
4. Sally re-enters: $a_4 = oc(S \triangleleft A \wedge A \triangleleft S)$

$s_4 = s_1 \otimes a_2 \otimes a_3 \otimes a_4 =$ 

We have $s_4 \models B_S \mathit{large}$. Thus again the modeller will pass the Sally-Anne test.

Modelling Sally-Anne in the new language

1. Sally has placed cube in large container: $s_1 = \begin{tikzcd} \bullet_{large, S\lhd A, A\lhd S} \arrow[loop above, "S, A"] \end{tikzcd}$
2. Sally leaves the room: $a_2 = oc(\neg S\lhd A \wedge \neg A\lhd S)$
3. Anne transfers cube: $a_3 = do(A, \neg large \wedge small)$
4. Sally re-enters: $a_4 = oc(S\lhd A \wedge A\lhd S)$

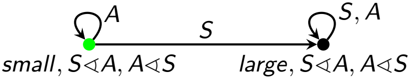
$$s_4 = s_1 \otimes a_2 \otimes a_3 \otimes a_4 = \begin{tikzcd} \underset{small, S\lhd A, A\lhd S}{\bullet} \arrow[loop above, "A"] \arrow[rr, "S"] & & \underset{large, S\lhd A, A\lhd S}{\bullet} \arrow[loop above, "S, A"] \end{tikzcd}$$

We have $s_4 \models B_S large$. Thus again the modeller will pass the Sally-Anne test.

But now we also have $s_1 \otimes a_3 = \begin{tikzcd} \bullet_{small, S\lhd A, A\lhd S} \arrow[loop above, "S, A"] \end{tikzcd} \neq s_4$. Hence our previous problem has been solved.

Modelling Sally-Anne in the new language

1. Sally has placed cube in large container: $s_1 = \overset{S,A}{\circlearrowleft} \; \mathit{large}, S \lhd A, A \lhd S$
2. Sally leaves the room: $a_2 = oc(\neg S \lhd A \wedge \neg A \lhd S)$
3. Anne transfers cube: $a_3 = do(A, \neg \mathit{large} \wedge \mathit{small})$
4. Sally re-enters: $a_4 = oc(S \lhd A \wedge A \lhd S)$

$s_4 = s_1 \otimes a_2 \otimes a_3 \otimes a_4 =$ 

We have $s_4 \models B_S \mathit{large}$. Thus again the modeller will pass the Sally-Anne test.

But now we also have $s_1 \otimes a_3 = \overset{S,A}{\circlearrowleft} \; \mathit{small}, S \lhd A, A \lhd S \neq s_4$. Hence our previous problem has been solved.

Full formalisation of Sally-Anne:
$do(A, \mathit{large}), oc(\neg S \lhd A \wedge \neg A \lhd S), do(A, \neg \mathit{large} \wedge \mathit{small}), oc(S \lhd A \wedge A \lhd S).$

Higher-order false-belief tasks

Full formalisation of second-order chocolate task:
$do(boy, drawer), oc(\neg boy \lhd girl \wedge \neg girl \lhd boy), oc(boy \lhd girl),$
$do(girl, \neg drawer \wedge box)$.

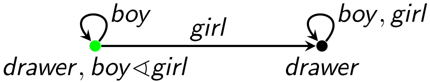
In resulting state s_4: $s_4 \models B_{girl} B_{boy} drawer$, as required.

Moreover, e.g.: $s_4 \models boy \lhd girl \wedge B_{girl} \neg boy \lhd girl \wedge B_{boy} B_{girl} \neg boy \lhd girl$.

Chocolate task in extended DEL versus stand. DEL

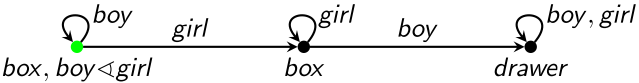
Epistemic model right before the girl moves the chocolate:

$s_3 =$



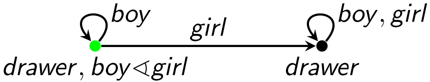
Applying the 2-event model $a_4 = do(girl, \neg drawer \wedge box)$ in s_3 we get:

$s_4 = s_3 \otimes a_4 =$

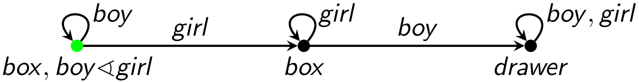


Chocolate task in extended DEL versus stand. DEL

Epistemic model right before the girl moves the chocolate:

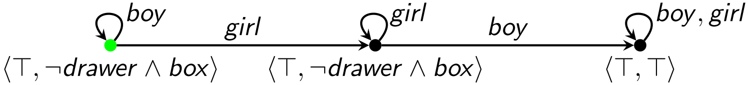
$s_3 =$ 

Applying the 2-event model $a_4 = do(girl, \neg drawer \wedge box)$ in s_3 we get:

$s_4 = s_3 \otimes a_4 =$ 

Proposition *Assume p is common belief in s, there is no nth order false-beliefs in s, and a is a* **standard** *2-event model. Then p can not be an nth-order false belief in $s \otimes a$. (simplified formulation)*

Hence the smallest standard event model that can produce s_4 from s_3 is this:



Robustness revisited

We have formalised the first-order *Sally-Anne task* and the second-order *chocolate task*.

For **robustness**, the formalism should be able to deal with tasks of **arbitrary order**. Proving this formally is future work.

Robustness revisited

We have formalised the first-order *Sally-Anne task* and the second-order *chocolate task*.

For **robustness**, the formalism should be able to deal with tasks of **arbitrary order**. Proving this formally is future work.

At least: We can go from a formalisation of a first-order task to a second-order task "at no extra cost", which is not possible in standard DEL.

Robustness revisited

We have formalised the first-order *Sally-Anne task* and the second-order *chocolate task*.

For **robustness**, the formalism should be able to deal with tasks of **arbitrary order**. Proving this formally is future work.

At least: We can go from a formalisation of a first-order task to a second-order task "at no extra cost", which is not possible in standard DEL.

Formalising other well-known false-belief tasks:

- **Ice-cream task** [Perner and Wimmer, 1985].
- **Birthday puppy task** [Sullivan et al., 1994].
- **Clown in the park task** [Wahl and Spada, 2000].

These all involve *untruthful announcements*. We need a more expressive framework: *plausibility models* [Baltag and Smets, 2008]. Future work.

Faithfulness revisited

A big step in the right direction:

$$\begin{array}{lcl} \text{agent } i \text{ makes } \phi \text{ true} & \curvearrowright & do(i, \phi) \\ i \text{ starts observing } j & \curvearrowright & oc(i \triangleleft j) \end{array}$$

Full formalisation of Sally-Anne:
$do(A, large), oc(\neg S \triangleleft A \wedge \neg A \triangleleft S), do(A, \neg large \wedge small), oc(S \triangleleft A \wedge A \triangleleft S)$.

Full formalisation of second-order chocolate task:
$do(boy, drawer), oc(\neg boy \triangleleft girl \wedge \neg girl \triangleleft boy), oc(boy \triangleleft girl)$,
$do(girl, \neg drawer \wedge box)$.

Summary

- I presented a formalisation of two false-belief tasks in an extension of DEL.

Summary

- I presented a formalisation of two false-belief tasks in an extension of DEL.
- The DEL extension is expected to have independent future interest (e.g. in epistemic planning).

Summary

- I presented a formalisation of two false-belief tasks in an extension of DEL.
- The DEL extension is expected to have independent future interest (e.g. in epistemic planning).

Current extensions of the presented work:

- From false-belief **analysis** to **synthesis**: planning to achieve false belief in others (deceit).

Summary

- I presented a formalisation of two false-belief tasks in an extension of DEL.
- The DEL extension is expected to have independent future interest (e.g. in epistemic planning).

Current extensions of the presented work:

- From false-belief **analysis** to **synthesis**: planning to achieve false belief in others (deceit).
- False-belief tasks using **plausibility models** and adding **abduction** (with Sonja Smets).

Summary

- I presented a formalisation of two false-belief tasks in an extension of DEL.
- The DEL extension is expected to have independent future interest (e.g. in epistemic planning).

Current extensions of the presented work:

- From false-belief **analysis** to **synthesis**: planning to achieve false belief in others (deceit).
- False-belief tasks using **plausibility models** and adding **abduction** (with Sonja Smets).
- Devising classes of false-belief tasks of **arbitrary order**, and prove them to be formalisable in the framework.

Summary

- I presented a formalisation of two false-belief tasks in an extension of DEL.
- The DEL extension is expected to have independent future interest (e.g. in epistemic planning).

Current extensions of the presented work:

- From false-belief **analysis** to **synthesis**: planning to achieve false belief in others (deceit).
- False-belief tasks using **plausibility models** and adding **abduction** (with Sonja Smets).
- Devising classes of false-belief tasks of **arbitrary order**, and prove them to be formalisable in the framework.
- Properties of edge-conditioned models: **exponential succinctness**, etc.

Appendix: Modelling choices for observations

What should observations be connected to? Several possibilities:

- **Propositions**. Proposition p is observed by agent i if ...
- **All actions**. All actions taking place are observed by agent i if ...
- **Particular actions**. Action a is observed by agent i if ...
- **All actions of particular agents**. The actions of agent j is observed by agent i if ...

	axiom encoded	**state encoded**
propositions	[Brenner and Nebel, 2009] sensor models Axioms: **sensor**$(i, p, cond)$	[Hoek et al., 2011] Note: observable propositions are **fixed**
all actions		[van Ditmarsch et al., 2013] New propositions: h_i means i *is paying attention*
particular actions	[Baral et al., 2012] Action language $m\mathcal{A}+$ Axioms: i **observes** a **if** ϕ	
Actions of agents		

Appendix: References I

Arkoudas, K. and Bringsjord, S. (2008).
Toward Formalizing Common-Sense Psychology: An Analysis of the False-Belief Task.
In PRICAI, (Ho, T. B. and Zhou, Z.-H., eds), vol. 5351, of Lecture Notes in Computer Science pp. 17–29, Springer.

Arslan, B., Taatgen, N. and Verbrugge, R. (2013).
Modeling Developmental Transitions in Reasoning about False Beliefs of Others.
In Proc. of the 12th International Conference on Cognitive Modelling.

Baltag, A., Moss, L. S. and Solecki, S. (1998).
The Logic of Public Announcements and Common Knowledge and Private Suspicions.
In Proceedings of the 7th Conference on Theoretical Aspects of Rationality and Knowledge (TARK-98), (Gilboa, I., ed.), pp. 43–56, Morgan Kaufmann.

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.

Baral, C., Gelfond, G., Pontelli, E. and Son, T. C. (2012).
An action language for reasoning about beliefs in multi-agent domains.
In Proceedings of the 14th International Workshop on Non-Monotonic Reasoning vol. 4,.

Baron-Cohen, S. (1997).
Mindblindness: An essay on autism and theory of mind.
MIT press.

Appendix: References II

Bolander, T. (2014).
Seeing is Believing: Formalising False-Belief Tasks in Dynamic Epistemic Logic.
In Proceedings of the European Conference on Social Intelligence (ECSI-2014), (Herzig, A. and Lorini, E., eds), vol. 1283, of CEUR Workshop Proceedings pp. 87–107, CEUR-WS.org.

Braüner, T. (2013).
Hybrid-logical reasoning in false-belief tasks.
In Proceedings of Fourteenth Conference on Theoretical Aspects of Rationality and Knowledge (TARK), (Schipper, B., ed.), pp. 186–195,.

Breazeal, C., Gray, J. and Berin, M. (2011).
Mindreading as a foundational skill for socially intelligent robots.
In Robotics Research pp. 383–394. Springer.

Brenner, M. and Nebel, B. (2009).
Continual planning and acting in dynamic multiagent environments.
Autonomous Agents and Multi-Agent Systems *19*, 297–331.

Hoek, W. v. d., Troquard, N. and Wooldridge, M. (2011).
Knowledge and control.
In The 10th International Conference on Autonomous Agents and Multiagent Systems–Volume 2 pp. 719–726, International Foundation for Autonomous Agents and Multiagent Systems.

Perner, J. and Wimmer, H. (1985).
"John thinks that Mary thinks that..." attribution of second-order beliefs by 5-to 10-year-old children.
Journal of experimental child psychology *39*, 437–471.

Appendix: References III

Seligman, J., Liu, F. and Girard, P. (2013).
Facebook and the epistemic logic of friendship.
In Proceedings of Fourteenth Conference on Theoretical Aspects of Rationality and Knowledge (TARK), (Schipper, B., ed.), pp. 229–238,.

Sindlar, M. P. (2011).
In the Eye of the Beholder: Explaining Behavior through Mental State Attribution.
PhD thesis, Universiteit Utrecht.

Sullivan, K., Zaitchik, D. and Tager-Flusberg, H. (1994).
Preschoolers can attribute second-order beliefs.
Developmental Psychology *30*, 395.

van Ditmarsch, H., Herzig, A., Lorini, E. and Schwarzentruber, F. (2013).
Listen to me! Public announcements to agents that pay attention—or not.
In Logic, Rationality, and Interaction pp. 96–109. Springer.

van Ditmarsch, H. and Kooi, B. (2008).
Semantic Results for Ontic and Epistemic Change.
In Logic and the Foundation of Game and Decision Theory (LOFT 7), (Bonanno, G., van der Hoek, W. and Wooldridge, M., eds), Texts in Logic and Games 3 pp. 87–117, Amsterdam University Press.

Wahl, S. and Spada, H. (2000).
Children's reasoning about intentions, beliefs and behaviour.
Cognitive Science Quarterly *1*, 3–32.

Wimmer, H. and Perner, J. (1983).
Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception.
Cognition *13*, 103–128.