Hybrid-Logical Reasoning in the Smarties and Sally-Anne Tasks

What Goes Wrong When Incorrect Responses are Given?

Braüner, Torben

Published in:
Proceedings of the 37th Annual Meeting of the Cognitive Science Society, Pasadena, California, USA

Publication date:
2015

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (APA):
Braüner, T. (2015). Hybrid-Logical Reasoning in the Smarties and Sally-Anne Tasks: What Goes Wrong When Incorrect Responses are Given? In *Proceedings of the 37th Annual Meeting of the Cognitive Science Society, Pasadena, California, USA* (pp. 273-278). Cognitive Science Society.

Hybrid-Logical Reasoning in the Smarties and Sally-Anne Tasks: What Goes Wrong When Incorrect Responses are Given?

Torben Braüner (torben@ruc.dk)
Roskilde University, Denmark

Abstract

The present paper is a follow-up to the journal paper (Braüner, 2014) which in turn is a revised and extended version of the conference paper (Braüner, 2013). These papers were concerned with formalizations of the reasoning when giving correct responses to the psychological tests called the Sally-Anne task and the Smarties task, testing children's capacity to ascribe false beliefs to others. In the present paper we give an analysis of what goes wrong when incorrect answers are given. Our analysis corroborates the claim that children under four and autistic children have difficulties shifting to a perspective different from their own.

Keywords: False-belief tasks; hybrid logic; natural deduction

Introduction

In the area of cognitive psychology there is a reasoning task called the *Sally-Anne task*. The following is one version.

> *A child is shown a scene with two doll protagonists, Sally and Anne, having respectively a basket and a box. Sally first places a marble into her basket. Then Sally leaves the scene, and in her absence, the marble is moved by Anne and hidden in her box. Then Sally returns, and the child is asked: "Where will Sally look for her marble?"*

It is well-known from experiments that most children above the age of four correctly respond where Sally must falsely believe the marble to be (in the basket) whereas younger children respond where they know the marble to be (in the box). For autistic children the cutoff age is higher than four years.

The Sally-Anne task is one out of a family of reasoning tasks called *false-belief tasks* showing the same pattern, that most children above four answer correctly, but autistic children have to be older. Many researchers in cognitive psychology have argued that there is a link between autism and a lack of what is called *theory of mind*, which is a capacity to ascribe mental states to oneself and to others, for example beliefs. For a very general formulation of the theory of mind deficit hypothesis of autism, see the book (Baron-Cohen, 1995).

Giving a correct answer to the Sally-Anne task involves a shift of perspective to another person, namely Sally. You have to put yourself in another person's shoes, so to speak.[1] Since the capacity to take another perspective is a precondition for figuring out the correct answer to the Sally-Anne task and other false-belief tasks, the fact that autistic children have a higher cutoff age is taken to support the claim that autists have a limited or delayed theory of mind.

[1]This phrase might suggest that we are adopting what is known as the simulation-theory view of theory of mind. This is a matter of on-going consideration for us, but at the present stage we use this terminology in a pre-theoretical sense, since it expresses an intuition that we are interested in modelling in formal logic.

In a range of works Michiel van Lambalgen and co-authors have given a detailed logical analysis (but not a full formalization) of the reasoning taking place in the Sally-Anne task and other false-belief tasks in terms of non-monotonic closed world reasoning as used in logic programming, see in particular the book (Stenning & van Lambalgen, 2008). The paper (Arkoudas & Bringsjord, 2008) describes how the reasoning in the Sally-Anne task has been implemented in an interactive theorem prover using axioms and proof-rules formulated in a many-sorted first-order modal logic. The proof-rules employed in (Stenning & van Lambalgen, 2008) and (Arkoudas & Bringsjord, 2008) do not explicitly formalize the perspective shift required to pass the Sally-Anne task.

In the papers (Braüner, 2014) and (Braüner, 2013) we gave a logical analysis of the perspective shift required to give correct answers to the Sally-Anne task and another false-belief task called the Smarties task, and we demonstrated that these tasks can be fully formalized in a hybrid-logical natural deduction system originally introduced by Jerry Seligman in the 1990s. Based on the formalizations of (Braüner, 2014) and (Braüner, 2013), in the present paper we give an analysis of what goes wrong when incorrect answers are given. In the following two sections we explain why a *natural deduction* system for *hybrid modal logic* is appropriate to analyse the reasoning in the Sally-Anne and Smarties tasks, reflecting the shift between different perspectives.

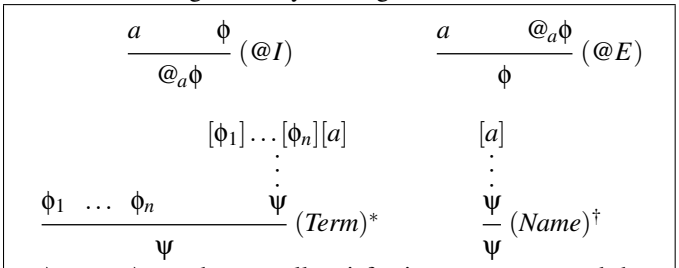
Since this paper is a follow-up to (Braüner, 2014), but space limitations only allows a brief recapitulation of the latter paper, the reader is advised to obtain a copy of that paper.

Hybrid modal logic

In the standard Kripke semantics for modal logic, the truth-value of a formula is relative to points in a set, that is, a formula is evaluated "locally" at a point, where points usually are taken to represent possible worlds, times, locations, persons, epistemic states, states in a computer, or something else. Hybrid logics are extended modal logics where it is possible to directly refer to such points in the logical object language, whereby locality can be handled explicitly.

The most basic hybrid logic is obtained by extending ordinary modal logic with *nominals*, which are propositional symbols of a new sort, each interpreted in a restricted way, being true at exactly one point. Most hybrid logics involve further additional machinery; here we shall consider a kind of operator called *satisfaction operators*. The motivation for adding satisfaction operators is to be able to formalize a statement being true at a particular time, location, or something else. In general, if a is a nominal and ϕ is an arbitrary for-

Figure 1: Hybrid-logical rules



∗ $\phi_1, \ldots, \phi_n$, and ψ are all satisfaction statements and there are no undischarged assumptions in the derivation of ψ besides the specified occurrences of $\phi_1, \ldots, \phi_n$, and a.
† a does not occur in ψ or in any undischarged assumptions other than the specified occurrences of a.

mula, then a new formula $@_a\phi$ can be built, where $@_a$ is a a satisfaction operator. The formula $@_a\phi$ expresses that the formula ϕ is true at one particular point, namely the point to which the nominal a refers. See the book (Braüner, 2011) for the formal syntax and semantics of hybrid logic.

When points in the Kripke semantics represent local perspectives (times or persons), hybrid logic can handle the different perspectives in the Sally-Anne and Smarties task.

Seligman's natural deduction system

Formal proofs built according to the rules of proof systems can be used to represent (describe the structure of) mathematical arguments as well as arguments in everyday human practice.

Natural deduction style proofs are meant to formalize the way human beings actually reason, and there is even experimental support for natural deduction being the mechanism underlying human deductive reasoning, (Rips, 2008). This is the main claim of the "mental logic" school in the psychology of reasoning (whose major competitor is the "mental models" school, claiming that the mechanism underlying human reasoning is the construction of models).

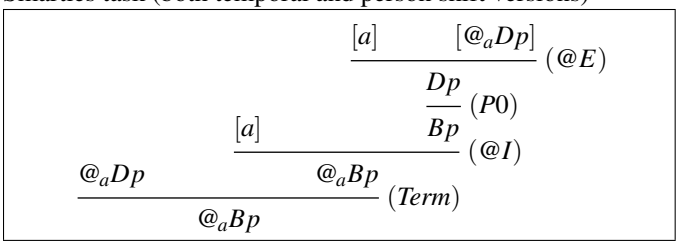
In general, natural deduction systems have two different kinds of rules for each connective; there are rules which introduce a connective and there are rules which eliminate a connective. Natural deduction rules may discharge assumptions which is indicated by putting brackets $[\ldots]$ around the assumptions in question.

Now, Seligman's natural deduction system is obtained by extending the standard natural deduction system for propositional logic with the rules in Figure 1 (we ignore modal operators since they are not relevant here). The system, which is a modified version of the system originally introduced by Jerry Seligman, is taken from Chapter 4 of (Braüner, 2011).

The rules $(@I)$ and $(@E)$ in Figure 1 are the introduction and elimination rules for the satisfaction operator.

The rule $(Term)$ in Figure 1 enables hypothetical reasoning where reasoning is about what is the case at a specific possible world (time or person), possibly different from the actual

Figure 2: Formalization of the child's correct response in the Smarties task (both temporal and person shift versions)



world. The hypothetical reasoning is formalized by the subderivation delimited by the rule, and the hypothetical world is the world referred to by the nominal discharged by the rule—indicated by $[a]$ in the $(Term)$ rule. This nominal might be called the point-of-view nominal. The $(Term)$ rule captures particularly well the perspective shift taking place when giving a correct answer to the Sally-Anne and Smarties tasks.

Correct response in the Smarties task

We start with a brief description of how the paper (Braüner, 2014) formalizes the correct reasoning in the Smarties task. The Smarties task comes in two versions, namely a version where there is a shift of perspective to an earlier time, and a version where there is a shift of perspective to another person. Here is the temporal version.

> *A child is shown a Smarties tube where unbeknownst to the child the Smarties have been replaced by pencils. The child is asked: "What do you think is inside the tube?" The child answers "Smarties!" The tube is then shown to contain pencils only. The child is then asked: "Before this tube was opened, what did you think was inside?"*

First an informal analysis. Let us call the child Peter. Let a be the time when Peter answers the first question, and t the time where he answers the second one. To answer the second question, Peter imagines himself being at the earlier time a where he was asked the first question. At that time he deduced that there were Smarties inside the tube from the fact that it is a Smarties tube. Imagining being at the time a, Peter reasons that since he at that time deduced that there were Smarties inside, he must also have come to believe that there were Smarties inside. Therefore, at t he concludes that at the earlier time a he believed that there were Smarties inside.

We now extend the language of hybrid logic with two modal operators, D and B. We make use of the following symbolizations

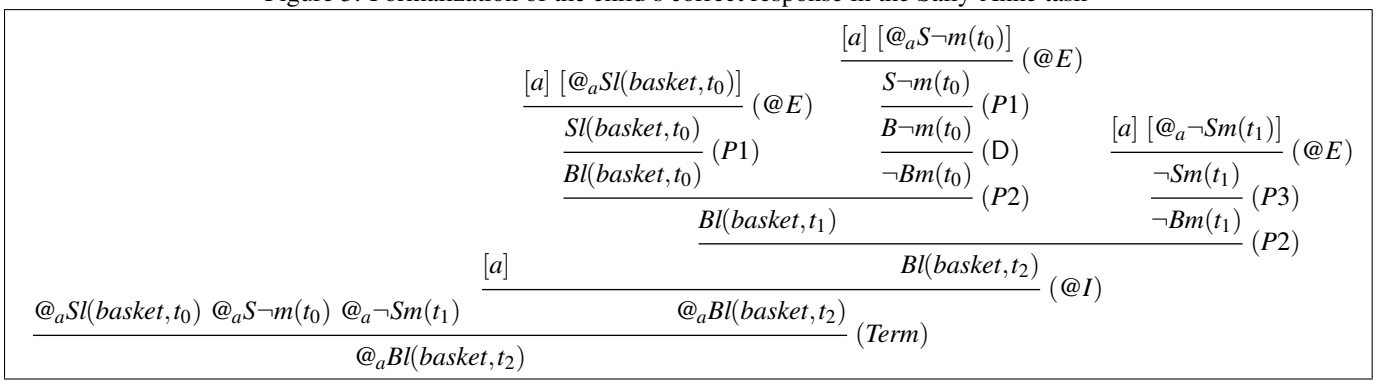
p	There are Smarties inside the tube
D	Peter deduces that ...
B	Peter believes that ...
a	The time where the first question is answered

and we take the principle

$(P0)$ $\quad D\phi \rightarrow B\phi$

as an axiom. This is principle (9.4) in (Stenning & van Lambalgen, 2008), page 251.

Figure 3: Formalization of the child's correct response in the Sally-Anne task



Then the shift of temporal perspective in the Smarties task can be formalized very directly as the derivation in Figure 2, where a is the point-of-view nominal, and where we have used a rule-version of the principle $(P0)$ above (more compact and more in the spirit of natural deduction). The premise $@_a Dp$ says that Peter at the earlier time a deduced that there were Smarties inside the tube, which he remembers at t.

We also take a look at the person version of the Smarties task. The only difference between the two versions is the second question where

> *"Before this tube was opened, what did you think was inside?"*

is replaced by

> *"If your mother comes into the room and we show this tube to her, what will she think is inside?"*

To give a correct answer to the latter of these two questions, the child Peter imagines being the mother coming into the room. Imagining being the mother, Peter reasons that the mother must deduce that there are Smarties inside the tube from the fact that it is a Smarties tube, and from that, she must also come to believe that there are Smarties inside. Therefore, Peter concludes that the mother would believe that there are Smarties inside.

The derivation formalizing this argument is exactly the same as in the temporal case, Figure 2, but some symbols are interpreted differently, namely

- D Deduces that ...
- B Believes that ...
- a The imagined mother

So now nominals refer to persons rather than times. Thus, the premise $@_a Dp$ in the derivation in Figure 2 says that the imagined mother deduces that there are Smarties inside the tube, which the child doing the reasoning takes to be the case since the mother is imagined to be present in the room.

Correct response in the Sally-Anne task

In this section we give a brief description of how (Braüner, 2014) formalizes the correct reasoning in the Sally-Anne task. Let us call the child Peter again. We shall consider three successive times t_0, t_1, t_2 where t_0 is the time at which Sally leaves the scene, t_1 is the time at which the marble is moved to the box, and t_2 is the time after Sally has returned when Peter answers the question. To answer the question, Peter imagines himself being Sally, and he reasons as follows: At the time t_0 when Sally leaves, she believes that the marble is in the basket since she sees it, and she sees no action to move it, so when she is away at t_1, she also believes the marble is in the basket. At t_2, after she has returned, she still believes that the marble is in the basket since she has not seen Anne moving it at the time t_1. Therefore, Peter concludes that Sally believes that the marble is in the basket.

In our formalization we make use of the predicates $l(i,t)$ and $m(t)$ as well as the modal operators S and B. The argument i in the predicate $l(i,t)$ denotes a location, and the argument t in $l(i,t)$ and $m(t)$ denotes a timepoint. We take time to be discrete, and the successor of t is denoted $t+1$.

- $l(i,t)$ The marble is at location i at time t
- $m(t)$ The marble is moved at time t
- S Sees that ...
- B Believes that ...
- a The person Sally

We also make use of the following four principles

- (D) $B\phi \to \neg B\neg\phi$
- $(P1)$ $S\phi \to B\phi$
- $(P2)$ $Bl(i,t) \wedge \neg Bm(t) \to Bl(i,t+1)$
- $(P3)$ $Bm(t) \to Sm(t)$

Principle (D) is a common modal axiom and it says that beliefs are consistent, that is, if something is believed, then its negation is not also believed. Strictly speaking, we use $B\neg\phi \to \neg B\phi$ which is equivalent to (D).

Principle $(P1)$ formalizes how a belief in something may be formed, namely by seeing it. This is principle (9.2) in the book (Stenning & van Lambalgen, 2008), page 251.

Principle $(P2)$ is remiscent of principle (9.11) in (Stenning & van Lambalgen, 2008), page 253, and axiom $[A_5]$ in (Arkoudas & Bringsjord, 2008), page 20. Principle $(P2)$ formalizes a "principle of inertia" saying that a belief in the predicate l being true is preserved over time, unless it is believed that an action has taken place causing the predicate to be false.

Principle $(P3)$ encodes the information that *seeing* the marble being moved is the only way a belief that the marble is

being moved can be acquired.

The shift of person perspective in the Sally-Anne task can now be formalized as the derivation in Figure 3, where a is the point-of-view nominal. The first two premises $@_aSl(basket,t_0)$ and $@_aS\neg m(t_0)$ say that Sally at the earlier time t_0 saw that the marble was in the basket and that no action was taken to move it, which the child Peter remembers. The third premise, $@_a\neg Sm(t_1)$, says that Sally did not see the marble being moved at the time t_1, this being the case since she was absent, which Peter remembers.

What goes wrong when incorrect responses are given?

The derivations given in Figure 2 and Figure 3 are formalizations of the reasoning taking place when correct answers are given to the Smarties and Sally-Anne tasks. The correct responses are summed up below in Table 1.

Table 1	Correct response	Formula
Smarties (temporal version)	At the time of question one Peter believes that the tube contains Smarties	$@_aBp$
Smarties (person version)	The imagined mother believes that the tube contains Smarties	$@_aBp$
Sally-Anne	Sally believes that the marble is in the basket at the time t_2	$@_aBl(basket,t_2)$

As can be seen from Figure 2 and Figure 3, the formulas in Table 1 are derived via a perspective shift to the point-of-view nominal a, standing for respectively the time where the first question is answered, the imagined mother, and the doll Sally.

Let b be the childs own perspective[2], that is, in the temporal version of the Smarties task, b is the time where the second question is answered, and in person version of the Smarties task, and in the Sally-Anne task as well, b is the person Peter. So to derive the correct answers in Figure 2 and Figure 3, there is a shift of perspective from b to a, and then back to b.

Now, the derivations of the correct answers in Figure 2 and Figure 3 do not explicitly tell what goes wrong when incorrect answers are given. But a child either answers correctly, or tends to give a specific incorrect answer: In case of the Smarties task, the child answers "Pencils", that is, the real content of the tube, not "Cereals" or something else irrelevant. Similarly, in the Sally-Anne task, the child reports the real location of the marble. Thus, there is a systematic tendency to report one's own belief, rather than that of another person—a phenomenon which we shall discuss in the next section. In what follows, we will analyze this pattern in the incorrect answers. To this end we let the propositional symbol q symbolize "The tube contains pencils". Then the incorrect answers can be summed up as follows.

[2]Note that b is not indicated in the formal derivations in Figure 2 and Figure 3, like it is not part of a formal mathematical proof that it has been carried out by a certain mathematician. The formal derivation itself does not care whether it is a certain human that carries out the reasoning, or the reasoning takes place in a computer, or in some other medium. Note also that b is actually indicated in Figure 4 and Figure 5, but this is because the latter derivations are about what is the case from the perspective b, which happens to be the perspective of the child carrying out the reasoning.

Figure 4: Formalization of the child's reasoning in the Smarties task (what is the case from its own perspective)

$$\dfrac{b \qquad \dfrac{\dfrac{b \qquad @_bSq}{Sq}\,(@E)}{Bq}\,(P1)}{@_bBq}\,(@I)$$

Table 2	Incorrect response	Formula
Smarties (temporal version)	At the time of question one Peter believes that the tube contains pencils	$@_aBq$
Smarties (person version)	The imagined mother believes that the tube contains pencils	$@_aBq$
Sally-Anne	Sally believes that the marble is in the box at the time t_2	$@_aBl(box,t_2)$

The three formulas in Table 2 are false in the scenarios described by the reasoning tasks, but here is an important observation: If we replace the perspective a in the formulas above by the childs own perspective b, then we obtain true formulas, namely the following.

Table 3	True proposition	Formula
Smarties (temporal version)	At the time of question two Peter believes that the tube contains pencils	$@_bBq$
Smarties (person version)	Peter believes that the tube contains pencils	$@_bBq$
Sally-Anne	Peter believes that the marble is in the box at the time t_2	$@_bBl(box,t_2)$

Below we demonstrate that the formulas $@_bBq$ and $@_bBl(box,t_2)$ in Table 3 are true by giving derivations in Seligman's system extended with the principles introduced in the previous section.

The formula $@_bBq$ in Table 3 can be derived from b and $@_bSq$ by the very simple derivation in Figure 4, where the nominal b is true since it is the perspective of the child who is doing the reasoning, and $@_bSq$ is obviously true in both the temporal and the person version, in both cases since the child when the second question is answered sees that there are pencils inside the tube. The formula $@_bBl(box,t_2)$ in Table 3 can be derived from b together with $@_bSl(box,t_1)$ and $@_bS\neg m(t_1)$ by the derivation in Figure 5. Again, the nominal b is true since it stands for the child Peter who happens to be the one doing the reasoning. The formulas $@_bSl(box,t_1)$

Figure 5: Formalization of the child's reasoning in the Sally-Anne task (what is the case from its own perspective)

$$\dfrac{\dfrac{\dfrac{b \quad @_b Sl(box,t_1)}{Sl(box,t_1)}\,(@E)}{Bl(box,t_1)}\,(P1) \qquad \dfrac{\dfrac{\dfrac{b \quad @_b S\neg m(t_1)}{S\neg m(t_1)}\,(@E)}{B\neg m(t_1)}\,(P1)}{\neg Bm(t_1)}\,(\mathsf{D})}{\dfrac{b \qquad Bl(box,t_2)}{@_b Bl(box,t_2)}\,(@I)}\,(P2)$$

and $@_b S\neg m(t_1)$ say that Peter at the earlier time t_1 saw that the marble was in the box and that no action was taken to move it, which Peter remembers. From these formulas, the formula $@_b Bl(box,t_2)$ is derived using the principle of inertia $(P1)$ and other principles. The principle of inertia is needed since Peter cannot see the content of the box at t_2, but at t_1 he came to believe that the marble was in the box, and this belief is preserved over time to t_2, since he does not believe that an action was taken to move the marble.

Note that the rule $(Term)$ is not used in the derivations in Figure 4 and Figure 5, and since b is the childs own perspective, there is no shift to a different perspective taking place in these derivations.

The formulas considered in the three tables above can be classified along the following two dimensions.

Table 4	The second perspective (the nominal a)	The childs own perspective (the nominal b)
Involves the false statements p and $l(basket,t_2)$	Correct responses cf. Table 1 $@_a Bp$ and $@_a Bl(basket,t_2)$	
Involves the true statements q and $l(box,t_2)$	Incorrect responses cf. Table 2 $@_a Bq$ and $@_a Bl(box,t_2)$	True statements cf. Table 3 $@_b Bq$ and $@_b Bl(box,t_2)$

The last table, Table 4, shows a pattern: The child giving an incorrect response (lower left quarter) reports what is believed to be the case from the childs own perspective (lower right quarter), and the child does not perform the shift of perspective required to be able to report what is believed to be the case from the second perspective (upper left quarter). Thus, this "pattern of failure" gives a formal corroboration of the claim that children under four and autistic children have difficulties shifting to a perspective different from their own.

Relation to realist bias

In Table 2, and the lower left quarter of Table 4, we summed up the incorrect responses to the Smarties and Sally-Anne tasks, where the subjects report their own belief, rather than that of another person, as required to give a correct answer. This systematic tendency to report what is believed to be true of reality, rather than what others might believe of reality, resembles the bias in adults' mindreading judgements which by some authors is called a *realist bias*, cf. (Mitchell, Robinson, Isaacs, & Nye, 1996), or *curse of knowledge*, cf. (Birch & Bloom, 2007). In the present setting, this realist bias, or curse of knowledge, amounts to reporting what is the case from the subject's own perspective, rather than what can be inferred to be the case from someone else's perspective.

The paper (Birch & Bloom, 2007) reports a study where the Sally-Anne scenario is extended such there are four containers instead of just two, and rather than judging where Sally would look, subjects rated the probability that she would look in each of the four containers. On some trials, the subjects knew where the marble really was, like in the original version of the Sally-Anne task where the subjects knew that the marble was in the box, but on other trials they only knew that it was in another container than initially. It turned out that when the subjects knew the real location of the marble, they judged it more likely that Sally would search in the real location, compared to when they did not know the real location.

The point above is that the subject's own knowledge about the real location is irrelevant—what matters is Sally's knowledge, which is the same in either case. In particular, note that whether or not the subject knows the actual location of the marble, this piece of information is obviously not included in Sally's knowledge, which is in line with the fact that the actual location of the marble, namely the box, is not even mentioned in the formalization of the correct response in Figure 3. Similarly, in the Smarties task, information about the actual content of the tube, namely pencils, is not involved in drawing the correct conclusion, that is, the propositional symbol q symbolizing "The tube contains pencils" is not mentioned in the formalization of the correct response in Figure 2.

The paper (Birch & Bloom, 2007) concerns adult subjects, but in the paper it is suggested that the difficulty children under four have on false-belief tasks should partially be accounted for in terms of an exaggerated curse-of-knowledge bias—not only in terms of conceptual limitations, that is, not only in terms of a limited concept of belief, or more generally, a limited concept of mental state, which is a common explanation in the literature.

The authors of (Birch & Bloom, 2007) in their earlier paper (Birch & Bloom, 2003) reported experiments involving three to five year old children, where it was demonstrated that three to four year old children were particularly susceptible to the curse-of-knowledge bias in comparison to five year old children. With reference to these earlier experiments, as well as other works, the paper (Birch & Bloom, 2007) calls for further experiments, where variants of false-belief tasks are used to clarify the role of the curse-of-knowledge bias in children's mental-state reasoning.

Where is the origin of mistakes?

As described earlier, the child giving an incorrect answer does not perform the shift of perspective required to figure out the

correct answer (upper left quarter in Table 4), but instead reports what is believed to be the case from the childs own perspective (lower right quarter in Table 4), namely the formulas $@_b Bq$ and $@_b Bl(box, t_2)$. But as shown in Figure 4 and Figure 5, these two formulas are actually derivable using hybrid-logical rules, thus, the incorrect answers can be derived using logically correct rules, that is, rules living up to a normative standard of logical correctness. This suggests that the origin of the mistakes lies in a wrong interpretation of the task, and not in the underlying logic[3].

This can be analyzed in terms of the two stages in reasoning emphasized in (Stenning & van Lambalgen, 2008), namely reasoning *to* and reasoning *from* an interpretation: First one fixes the domain of discourse and the interpretation of logical and non-logical expressions, and only after this has been achieved, a set of normatively correct formal rules can be determined, guiding one's reasoning. In terms of this distinction, the origin of the mistakes made by young children and autists seems to be located in the first stage, that is, in the reasoning to an interpretation of the task, rather than in the second stage.

According to (Stenning & van Lambalgen, 2008), page 25, the technical part of reasoning to an interpretation involves

i) fixing a formal language,

ii) fixing a semantics for the formal language, and

iii) fixing a definition of valid arguments in the language.

The semantics includes a notion of a mathematical representation of the domain, what we call a model, together with a definition of satisfaction, connecting the formal language to the mathematical models. In these technical terms, it seems plausible that the origin of the mistakes made by young children and autists lies in fixing a semantics, more specifically a Kripke model including only one perspective, namely the subject's own perspective.

Related work

The approach taken in the present work, based on (Braüner, 2013, 2014), is to model the reasoning in false-belief tasks from perspective of the subject doing the reasoning. Another approach is to use dynamic epistemic logic to model the reasoning from a global perspective, that is, from the perspective of the modeler, see for example (Bolander, 2014).

The paper (van Ditmarsch & Labuschagne, 2007) models examples of beliefs that agents may have about other agents' beliefs, one example is an autistic agent that always believes that other agents have the same beliefs as the agent's own. This is modelled by different agents preference relations between states, where an agent prefers one state over another if the agent considers it more likely. These beliefs turn out to be frame-characterizable by formulas of epistemic logic.

There are also a number of computational cognitive models of false-belief tasks, a recent example is (Arslan, Taatgen, & Verbrugge, 2013), which models the gradual development in false-belief reasoning using the so-called ACT-R cognitive architecture.

[3]Thanks to one of the anonymous reviewers for pointing this out.

Acknowledgements

Thanks to the anonymous reviewers for constructive feedback. The author acknowledges the funding received from The Velux Foundation for the project *Hybrid-Logical Proofs at Work in Cognitive Psychology* (VELUX 33305).

References

Arkoudas, K., & Bringsjord, S. (2008). Toward formalizing common-sense psychology: An analysis of the false-belief task. In T.-B. Ho & Z.-H. Zhou (Eds.), *PRICAI 2008: Trends in artificial intelligence* (Vol. 5351, pp. 17–29). Springer-Verlag.

Arslan, B., Taatgen, N., & Verbrugge, R. (2013). Modeling developmental transitions in reasoning about false beliefs of others. In *Proceedings of the 12th international conference on cognitive modeling* (pp. 77–82). Ottawa: Carleton University.

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. MIT Press.

Birch, S., & Bloom, P. (2003). Children are cursed: An asymmetric bias in mental state attribution. *Psychological Science*, *14*, 283–286.

Birch, S., & Bloom, P. (2007). The curse of knowledge in reasoning about false beliefs. *Psychological Science*, *18*, 382–386.

Bolander, T. (2014). Seeing is believing: Formalising false-belief tasks in dynamic epistemic logic. In A. Herzig & E. Lorini (Eds.), *Proceedings of the European conference on social intelligence (ECSI-2014)* (pp. 87–107). IRIT-CNRS, Toulouse University, France.

Braüner, T. (2011). *Hybrid logic and its proof-theory* (Vol. 37). Springer.

Braüner, T. (2013). Hybrid-logical reasoning in false-belief tasks. In B. Schipper (Ed.), *Proceedings of fourteenth conference on theoretical aspects of rationality and knowledge (TARK)* (pp. 186–195). (ISBN 978-0-615-74716-3, available at http://tark.org)

Braüner, T. (2014). Hybrid-logical reasoning in the Smarties and Sally-Anne tasks. *Journal of Logic, Language and Information*, *23*, 415–439. (Revised and extended version of (Braüner, 2013))

Mitchell, P., Robinson, E., Isaacs, J., & Nye, R. (1996). Contamination in reasoning about false belief: an instance of realist bias in adults but not children. *Cognition*, *59*, 1–21.

Rips, L. (2008). Logical approaches to human deductive reasoning. In J. Adler & L. Rips (Eds.), *Reasoning: Studies of human inference and its foundations* (pp. 187–205). Cambridge University Press.

Stenning, K., & van Lambalgen, M. (2008). *Human reasoning and cognitive science*. MIT Press.

van Ditmarsch, H., & Labuschagne, W. (2007). My beliefs about your beliefs – a case study in theory of mind and epistemic logic. *Synthese*, *155*, 191–209.