

Second-order false-beliefs, language and logic

Blackburn, Patrick Rowan; Braüner, Torben; Polyanskaya, Irina

Publication date:
2015

Document Version
Peer reviewed version

Citation for published version (APA):
Blackburn, P. R., Braüner, T., & Polyanskaya, I. (2015). *Second-order false-beliefs, language and logic*. Paper presented at Conference on Computing Natural Reasoning, Bloomington, United States.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the public portal.

Take down policy

If you believe that this document breaches copyright please contact rucforsk@kb.dk providing details, and we will remove access to the work immediately and investigate your claim.

Second-order false-beliefs, language and logic

Patrick Blackburn, Torben Braüner and Irina Polyanskaya
Roskilde University, Denmark

April 23, 2015

1 First-order false-belief tests

In cognitive psychology there is a reasoning task called the *Sally-Anne task*. Here is one version:

A child is shown a scene with two doll protagonists, Sally and Anne. Sally has a basket and Anne has 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?”

As is well-known from repeated experimentation, most typically developing children above the age of four correctly respond with the location where Sally (falsely) believes the marble to be (in the basket) whereas younger children respond with where they know the marble really is (in the box). For autistic children, on the other hand, the cutoff age is usually much higher than four years old, something that was first observed in [4].

The Sally-Anne task is simply one of a family of reasoning tasks called *first-order false-belief tasks*. Another such task is the Smarties task, and tasks in the first-order family lead to much the same result: typically developing children above the age of four usually answer them correctly, whereas autistic children are usually successful only when they are older. Tasks like the Sally-Anne and Smarties tasks are called *first-order false-belief tasks* since they measure a subject’s capacity to take into account other peoples beliefs about simple world facts. A *second-order false-belief task*, on the other hand, measures a subject’s capacity to take into account other people’s beliefs about other people’s beliefs (including their beliefs about the agent’s beliefs). In the next section we turn to second-order false-belief tasks, the topic we wish to explore in this paper.

Starting with [4], many researchers in cognitive psychology have argued that there is a link between autism and a lack of what is called *theory of mind* (ToM). This is a person’s capacity to ascribe mental states (such as beliefs) both to themselves and to others; for a very general formulation of the theory of mind deficit hypothesis of autism, see the book [3]. The results of first-order false-belief tasks are robust under many different variations, for example across various countries and various task manipulations, as shown in the meta-analysis [25], involving 178 individual false-belief studies and more than 4000 children.

Giving a correct answer to the Sally-Anne task involves a shift of perspective to another person (or, indeed, a doll), namely Sally. You have to put yourself in another agent’s shoes, so to speak. Our ‘shift of perspective’ terminology might suggest to a cognitive psychologist that we are adopting what is known as the *simulation approach* to theory of mind. But this is not our intention. Rather we are using this phrase in a pre-theoretical or intuitive sense, for it expresses an intuition that we are interested in modelling in formal logic.

Logical analysis of first-order false-belief tasks is not new. In a range of works Michiel van Lambalgen and various co-authors, notably Keith Stenning, have given a detailed logical analysis 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 [22].

Another formalization of the Sally-Anne task can be found in the paper [1] by Arkoudas and Bringsjord, one of their aims being “... to provide a formal model of false-belief attributions, and,

in particular, a description of the logical competence of an agent capable of passing a false-belief task” cf. p. 18. In their paper, Arkoudas and Bringsjord specifies axioms and proof-rules in many-sorted first-order modal logic, and use this machinery to implement the reasoning in the Sally-Anne test in an interactive theorem prover with a classical logic base.

The papers [7] and [6] by the second author of the present paper give 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 demonstrate that these tasks can be formalized in a hybrid-logical natural deduction system. Hybrid logics are extended modal logics where the object language allows direct reference to points in the Kripke model. In false-belief applications we take the points in the model to be the agents involved in the experiment. Using hybrid-logical machinery, the perspectives of individual persons can be handled explicitly, and, crucially, one can formulate statements about what is the case from the perspective of a specific person.

One way to compare the three formalizations of the Sally-Anne task just mentioned is in terms of the expressive power of the logics deployed, for example the amount of first-order machinery. Thus can be summed up as follows (see [7] for a more detailed comparison):

	The paper [7]	The book [22]	The paper [1]
Terms referring to times	Yes	Yes	Yes
Quantification over times	No	Yes	Yes
Quantification over events	No	No	Yes
Quantification over fluents	No	No	Yes

Another way to compare the formalizations is in terms of the logical principles used, for example whether logical omniscience¹ is implied. This is not the case with the belief modality in [6], but [22] does make use of principles implying logical omniscience. On the other hand, the axioms and proof-rules of [1] are tailor-made to avoid logical omniscience of the belief operator.

It should also be mentioned that there is newer work that formalizes false-belief tasks using dynamic epistemic logic, see for example [5]. This work models the reasoning from a global perspective, that is, from the perspective of the modeler. By way of contrast the three logical models discussed above all take the perspective of the subject doing the reasoning, which is the approach we will take. So we won’t attempt a precise comparison here.

2 Second-order false-belief tests

Second-order false-belief is an important topic, but less is known about it. There are a few different versions of second-order false-belief tasks, but far fewer than the huge number of first-order false-belief tasks. Below is an abridged version of a second-order false-belief task from the original paper on the topic [20].

A child is told a story about John and Mary. They play in the park where they see an ice-cream truck. John and Mary split up, and later, they are both told that the ice-cream truck has gone to the church, but neither knows that the other has been informed. The child is then asked: “Where does John think that Mary will go to buy ice cream?” and “Where does John think that Mary thinks the ice-cream truck is?”

So to pass a second-order false-belief task, the child has to realize that someone (John) can hold a false-belief about someone’s (Mary’s) belief about a state of affair in the world, in comparison to the first-order case, where the child “just” has to realize that someone (Sally) can hold a false-belief about a state of affair in the world. Consequently, second-order tasks are passed later than first-order tasks, typically between the ages of five to seven.

It is still unclear what causes the gap in age in second-order as opposed to first-order competence in false-belief tasks. As the review paper [17] explains, one position, the *conceptual change*

¹Logical omniscience says that knowledge is closed under logical consequence, that is, $K\psi$ can be derived from $\phi \rightarrow \psi$ and $K\phi$, which at least for human agents is implausible.

position, suggests that the move from first-order to second-order reasoning is difficult mainly because it involves conceptual enrichments (presumably including the realisation that other people can have beliefs about beliefs). But some argue that no conceptual change is involved, and that the increase in difficulty merely reflects the increased cognitive load that these more complex tasks require (this sometimes called the *complexity-only* position). In a well-known paper [23] Sullivan et al. introduced the puppy story, a simplification of the ice-cream story. This experiment gave rise to the narrowest gap in ages ever reported between the acquisition of first-order and second-order competence, and accordingly, the authors of [23] argued that the Puppy story showed that it is only the data processing load, and not the conceptually new understanding that is required for a child to pass it. To sum up, the conceptual change and complexity-only positions are empirically built on two stories: ice cream and puppy.

There are a limited number of publications dealing with formalizing second-order false-belief tests. Neither [1] nor [22] deal with second-order false-belief tests. The work of [7] by the second author of the present paper can be extended to encompass second-order tests, but this has not been published. But there are some higher-order logical formalizations along different lines: The paper [16] compares subjects' performance in the logically equivalent two-player games called Marble Drop and the Matrix Game, used to investigate higher-order social reasoning. The recent paper [5] formalizes the second-order *Chocolate task*² in dynamic epistemic logic.

3 What are logical analyses good for?

There can be a number of reasons for being interested in using logic to analyze and formalize psychological reasoning tasks.

We are primarily interested in using logic to compare reasoning tasks. One important case is where two tasks are superficially dissimilar, but turn out to have the same underlying logical structure. This is what was encountered in the paper [7] where it was shown that two different versions of the Smarties task (another person than the subject / the subject at another time) have exactly the same underlying logical structure. Other examples of dissimilar, but logically equivalent, reasoning tasks are the earlier mentioned games Marble Drop and the Matrix Game, which are game-theoretically equivalent, cf. [16]. If such investigations are based on proof-theory, they might be assisted by a notion of identity on proofs (exploiting the longstanding effort in proof-theory to give a notion of identity between proofs, that is, a way to determine when two arguments have common logical structure, despite superficial dissimilarity).

If two experiments make use of superficially dissimilar reasoning tasks, but which have the same underlying logical structure, then we would expect similar empirical results (for example in terms of number of correct answers and/or reaction time). In this case the identity of logical structure can be seen as an explanation of the similarity of the results. On the other hand, if the experiments give differing empirical results, despite having the same logical structure, then it calls for an explanation: One such explanation could be differing levels of abstraction, in the extreme case a purely symbolic reasoning task in comparison to a reasoning task dealing with a familiar everyday situation. An example is the above mentioned games Marble Drop and the Matrix Game where subjects perform better when a game is embedded in a concrete physical context (Marble Drop) than when it is given a more abstract formulation (the Matrix Game), as demonstrated in [16]. It is well-known from the literature that differences in the level of abstraction can give rise to differences in performance, see for example also the extensive literature on the *Wason selection task*, as surveyed in [22] and [19].

But the reverse can also happen. Sometimes one can detect that two similar tasks actually have different underlying logical structure. One example is the Sally-Anne task in comparison to the Smarties task: All three formalizations [7], [22] and [1] of the Sally-Anne task make use of a 'principle of inertia' saying that a belief is preserved over time, unless there is belief to the contrary, but neither [7] nor [22] make use of such a principle in the Smarties task (which is not considered

²Like the Sally-Anne task, but Sally cheats and peeks through the keyhole when she is out. Second-order question: "Where does Anne think that Sally think the marble is?"

in [1]). It might be investigated whether the theoretical observation that the Sally-Anne task depends on an inertia principle, but the Smarties task does not, can be detected empirically in terms of a concept widely applied in psychology, namely the Piagetian conservation concept, cf. [21]. This is a concept that children typically acquire at 5-7 years old, and to conserve in this context (Piaget’s terminology) is to preserve internally or represent.

Yet another reason for being interested in logical analyses of psychological reasoning tasks, is to try to find out what goes wrong when a subject gives an incorrect response, like the book [22] analyze the incorrect answers given to false-belief tasks by children under four and autistic children. See also the paper [8], which is a follow-up to [7]. Whereas the latter paper is concerned with hybrid-logical formalizations of the reasoning when giving correct answers to false-belief tests, the paper [8] gives an analysis of what goes wrong when incorrect responses are given—an analysis that corroborates the claim that children under four and autistic children have difficulties shifting to a perspective different from their own.

4 Relation between false-beliefs and language

The meta-analysis [18] reports several significant correlations between results on first-order false-belief tests and language tests, in particular what are called *memory for sentential complements tests*, like the following.

A child is told a story about Tom, who thinks that it is sunny outside - although it is really raining. The child is then asked: “Will Tom now put his raincoat on?” and “What was Tom thinking?”

A number of papers have investigated the more precise relationship between false-beliefs and mastery of sentential complements: The longitudinal study [9] suggests that mastery of sentential complements is a precursor of false-belief understanding, and the training study [15] suggests that sentential complement mastery actually plays a causal role in false-beliefs, which is corroborated by [13] and other papers as well.

There is not much work on second-order false-belief understanding and language, and it is not clear how the table below should be completed; we present our suggestions in the following section.

	related to
First-order false-belief tests	memory for complements tests
Second-order false-belief tests	???

A few papers on second-order false-beliefs and language can be mentioned: The paper [12] tests children on second-order false-belief tests, strategic games and a Dutch sentence comprehension test whereas [14] compares childrens performance on second-order false-belief tests and test of comprehension and production of multiple language embedding. The recent paper [2] tests children on second-order false-belief tests, double-embedded relative clause tests and other tests.

5 On-going work

We plan to test and analyse (monolingual, autistic, Danish native speaker) children’s performance on second-order false-beliefs and language. More precisely, we have a between-subjects empirical design that includes data collection and data analysis in two steps: one measuring correlations effect between variables, the other establishing causality effect via training. Three of the central inclusion criteria for the participants are that children of age 7-12 years have clinically established and confirmed Autism Spectrum Disorder, their IQ is within a normal range and there are no significant impairments in subject’s language development. As confounding variables we have taken verbal comprehension, working memory, grammar comprehension and use of expressive language.

As we have already mentioned, the review paper [17] notes that there are two different explanations of the time lag between first- and second-order mental understanding:

Conceptual change: A qualitative transformation of the underlying thought system, in particular, an understanding that mental states can be recursively embedded is acquired.

Information-processing complexity-only: Including linguistic capacity, the ability to track sequences of story information, reasoning through long inferential chains and memory load.

Accordingly, we think there are essentially two ways the table in the previous section can be completed, implying that there are two lines of work when considering Danish language tests that potentially correlate with, or are even causally related to, second-order false-belief tests. In the following two subsections we sketch two sorts of language tests, which we preliminarily hypothesize are related to respectively the complexity-only and conceptual change positions.

The JDV-test

One interesting Danish language test is the so-called JDV-test, [10], developed by Ditte Boeg Thomsen and Elisabeth Engberg-Pedersen, University of Copenhagen, and used to test typically developing adults and children as well as autistic children, [24, 11]. This test, which is the only existing Danish language test involving second-order mental state ascription, is a gap-filling test where the subject has to fill in one of the three Danish dialogue particles *jo*, *da* and *vel* (which have given the test its name). To choose the appropriate dialogue particle, the subject has to understand a story character’s understanding of another story character’s perspective. To quote from the paper [11] by Engberg-Pedersen and Boeg Thomsen.

Acquiring dialogue particles requires sophisticated perspective-taking skills as children must be able to entertain a state of affairs taking into account both their own mental state and another persons mental state simultaneously. ([11], page 1)

Thus, perspective-taking skills plays a key role in the JDV-test, like they arguably do in false-belief tests, cf. the logical analyses and formalizations given in [7, 8]. In the JDV-test, the prerequisite perspective-taking skills is reflected in the requirement that the subject has to understand the differences in meaning of the particles *jo*, *da* and *vel*.

- *Jo* means ”you know and I know, and we agree.”
- *Da* means ”you know and I know, but you (or somebody else) appear to disagree.”
- *Vel* means ”I believe, but you probably know better.”

In terms of knowledge, this requires understanding of the speaker’s own knowledge, others’ knowledge and shared knowledge. A sample item from the JDV-test, with English translation in red, is shown in Figure 1. The correct particle for this item is *jo* since both Jacob and Peter know that it is nice weather, and moreover, this is shared knowledge. We are particularly interested in investigating what ”shared” knowledge might mean, and we believe that this can be analyzed in terms of modal logic, where there is a clear-cut formal distinction between:

- An agent’s knowledge, modelled by a modal operator, $K_a\phi$
- An agent’s knowledge of another agent’s knowledge, $K_aK_b\phi$
- Common knowledge $C\phi$ between agents, which is equivalent to the conjunction of the countable set of formulas

$$\begin{aligned} & K_a\phi \wedge K_b\phi \\ & K_aK_a\phi \wedge K_aK_b\phi \wedge K_bK_a\phi \wedge K_bK_b\phi \\ & \vdots \end{aligned}$$

The modal operator C is in the Kripke semantics interpreted as the transitive closure of the union of the agent’s accessibility relations



Figure 1: Sample item from the JDV-test

Test involving recursion

In the present subsection we describe some test material designed to give another explanation of the time lag between first- and second-order mental understanding in mind. The underlying research question here is: do particular developments in syntax such as recursion comprehension affect particular kinds of thinking such as second-order ToM reasoning in autistic children? More concretely: does children's ability to understand (possibly also to produce) the linguistic structure involving or consisting of recursive structures predict performance on second-order false-belief tests, beyond development in general language abilities? In case of a positive answer, which logical, syntactic and semantic properties of recursive embedding can be hypothesized to be linked to false-belief understanding?

Here are two examples of test material we plan to use here:

Task: Jens talks to Eva. Mother is cleaning in the kitchen.
 Jens says to Eva that mother said that Spiderman was boring.
 Task question 1: What did Jens say to Eva?
 Task question 2: Who said that Spiderman was boring?

To be accompanied by two pictures: a boy talking with a girl, and a woman cleaning in the kitchen. And also the following.

Task: Pictures with a boy looking out of the window and it is raining outside (1),
 sunshine on the beach (2), a woman talking to the boy (3).
 Leo remembers/thinks that his mother said that the sun is shining outside, and that
 they're going to the beach.
 Test question: What does Leo remember/think?

6 Acknowledgements

The authors acknowledge the financial support received from the project *Hybrid-Logical Proofs at Work in Cognitive Psychology*, funded by The Velux Foundation (VELUX 33305).

References

- [1] K. Arkoudas and S. Bringsjord. Toward formalizing common-sense psychology: An analysis of the false-belief task. In T.-B. Ho and Z.-H. Zhou, editors, *PRICAI 2008: Trends in Artificial Intelligence*, volume 5351 of *Lecture Notes in Computer Science*, pages 17–29. Springer-Verlag, 2008.
- [2] B. Arslan, A. Hohenberger, and R. Verbrugge. The development of second-order social cognition and its relation with complex language understanding and memory. In *Proceedings of the 34th Annual Conference of the Cognitive Science Society*, pages 1290–1295. Cognitive Science Society, 2012.
- [3] S. Baron-Cohen. *Mindblindness: An Essay on Autism and Theory of Mind*. MIT Press, 1995.
- [4] S. Baron-Cohen, A.M. Leslie, and U. Frith. Does the autistic child have a 'theory of mind'? *Cognition*, 21:37–46, 1985.
- [5] T. Bolander. Seeing is believing: Formalising false-belief tasks in dynamic epistemic logic. In A. Herzig and E. Lorini, editors, *Proceedings of the European Conference on Social Intelligence (ECSI-2014)*, pages 87–107. IRIT-CNRS, Toulouse University, France, 2014.
- [6] T. Braüner. Hybrid-logical reasoning in false-belief tasks. In B.C. Schipper, editor, *Proceedings of Fourteenth Conference on Theoretical Aspects of Rationality and Knowledge (TARK)*, pages 186–195, 2013. ISBN 978-0-615-74716-3, available at <http://tark.org>.
- [7] T. Braüner. Hybrid-logical reasoning in the Smarties and Sally-Anne tasks. *Journal of Logic, Language and Information*, 23:415–439, 2014. Revised and extended version of [6].
- [8] T. Braüner. Hybrid-logical reasoning in the Smarties and Sally-Anne tasks: What goes wrong when incorrect responses are given? In *Proceedings of the 37th Annual Cognitive Science Society Meeting*. Pasadena, California: Cognitive Science Society, 2015. To appear.
- [9] J.G. de Villiers and J.E. Pyers. Complements to cognition: A longitudinal study of the relationship between complex syntax and false-belief-understanding. *Cognitive Development*, 17:1037–1060, 2002.
- [10] E. Engberg-Pedersen and D.B. Thomsen. JDV-test. University of Copenhagen, 2012.
- [11] E. Engberg-Pedersen and D.B. Thomsen. The socio-cognitive foundation of Danish perspective-mixing dialogue particles. In B. Dancygier, W.-l. Lu, and A. Verhagen, editors, *Linguistic Manifestations of Mixed Points of View in Narratives*. de Gruyter Mouton, 2015. In press.
- [12] L. Flobbe, R. Verbrugge, P. Hendriks, and I. Krämer. Children’s application of theory of mind in reasoning and language. *Journal of Logic, Language and Information*, 17:417–442, 2008.
- [13] C.H. Hale and H. Tager-Flusberg. The influence of language on theory of mind: a training study. *Developmental Science*, 6:346–359, 2003.
- [14] B. Hollebrandse, K. Hobbs, J. de Villiers, and T. Roeper. Second order embedding and second order false belief. In A. Gavarró and M.J. Freitas, editors, *Language Acquisition and Development, proceedings of GALA 2007*, pages 199–204. Cambridge Scholar Press, 2008.
- [15] H. Lohmann and M. Tomasello. The role of language in the development of false belief understanding: A training study. *Child Development*, 74:1130–1144, 2003.

- [16] B. Meijering, L. van Maanen, H. van Rijn, and R. Verbrugge. The facilitative effect of context on second-order social reasoning. In R. Catrambone and S. Ohlsson, editors, *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*, pages 1423–1428. Cognitive Science Society, 2010.
- [17] S.A. Miller. Children’s understanding of second-order mental states. *Psychological Bulletin*, 135:749–773, 2009.
- [18] K. Milligan, J.W. Astington, and L.A. Dack. Language and theory of mind: Meta-analysis of the relation between language ability and false-belief understanding. *Child Development*, 78:622–646, 2007.
- [19] C. Dutilh Novaes. *Formal Languages in Logic: A Philosophical and Cognitive Analysis*. Cambridge University Press, 2012.
- [20] J. Perner and H. Wimmer. “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, 1985.
- [21] J. Piaget. Quantification, conservation, and nativism. *Science*, 162:976–979, 1968.
- [22] K. Stenning and M. van Lambalgen. *Human Reasoning and Cognitive Science*. MIT Press, 2008.
- [23] K. Sullivan, D. Zaitchik, and H. Tager-Flusberg. Preschoolers can attribute second-order beliefs. *Developmental Psychology*, 30:395–402, 1994.
- [24] D.B. Thomsen. Viet til vinkler, viernes vinkler: typisk udviklede danske børns beherskelse af synsvinkelmarkerende diskurspartikler. University of Copenhagen, 2012. Kandidatspeciale.
- [25] H.M. Wellman, D. Cross, and J. Watson. Meta-analysis of theory-of-mind development: The truth about false-belief. *Child Development*, 72:655–684, 2001.