Second-order false-belief tasks

Analysis and formalization

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

Published in:
Logic, Language, Information, and Computation

DOI:
10.1007/978-3-662-52921-8_9

Publication date:
2016

Document Version
Peer reviewed version

Citation for published version (APA):
Braüner, T., Blackburn, P. R., & Polyanskaya, I. (2016). Second-order false-belief tasks: Analysis and formalization. In J. Väänänen, Å. Hirvonen, & R. de Queiroz (Eds.), *Logic, Language, Information, and Computation: 23rd International Workshop, WoLLIC 2016, Puebla, Mexico, August 16-19th, 2016. Proceedings* (pp. 125-144). Springer. Lecture Notes in Computer Science Vol. 9803 https://doi.org/10.1007/978-3-662-52921-8_9

Second-order false-belief tasks: Analysis and formalization*

Torben Braüner, Patrick Blackburn, and Irina Polyanskaya
Roskilde University, Denmark
{torben,patrickb,irinap}@ruc.dk

Abstract

We first give a coarse-grained modal-logical analysis of the four best known second-order false-belief tasks. This preliminary analysis shows that the four tasks share a common logical structure in which a crucial role is played by a "principle of inertia" which says that an agent's belief is preserved over time unless the agent gets information to the contrary. It also reveals informational symmetries (all four possibilities inherent in the two dimensions of deception versus no-deception and change-in-world versus change-in-belief-only are realized) and reveals a rather puzzling feature common to all four tasks. We then take a closer look at how the principle of inertia is used, which leads to a fine-grained analysis in terms of perspective shifting. We formalize this analysis using a natural deduction system for hybrid logic, and show that the proof modelling the solution to the first-order Sally-Anne task is nested inside the proof modelling the second-order solution.

1 Introduction

In this paper we use modal and hybrid logic to analyse four second-order false-belief tasks. We begin with our running example, the second-order Sally-Anne task (which was introduced in [3]):

> *A child is shown a scene with two doll protagonists, Sally and Anne, with a basket and a box respectively. Sally first places a marble into her basket. Then Sally leaves the scene, and in her absence, Anne moves the marble and puts it in her box.* ***However, although Anne does not realise this, Sally is peeking through the keyhole and sees what Anne is doing****. Then Sally returns, and the child is asked: "Where* ***does Anne think that*** *[Sally will] look for her marble?"*

Experiments have shown that typically developing children above the age of six usually handle second-order tasks correctly; see [13, 14]. They answer that Anne thinks that Sally will look in the basket, which is where Anne (falsely) believes that Sally believes the marble to be. Younger children usually answer that Anne thinks that Sally will look in the box: this is indeed where Sally knows the marble to be, but Anne does not know that Sally knows this, and hence the response is incorrect. In short, to pass the test, the experimental subject must ascribe a *false* belief to Anne, thus ensuring that the answer can't be explained as the subject simply reporting what is true — it really is Anne's *belief* that is being reported. For children with *Autism Spectrum Disorder (ASD)*, the shift to correct responses tends to occur at a later age, if it happens at all.

If the bold font material is deleted, and [Sally will] is switched to 'will Sally', our statement of the second-order Sally-Anne task becomes a statement of the well known first-order Sally-Anne task (it was introduced in [2]):

*The present paper is a post-print version of the paper [9]. The final publication is available at Springer via http://dx.doi.org/10.1007/978-3-662-52921-8_9

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

Extensive experimental work with first-order false-belief tasks has shown the existence of a transition age, and it is lower than in the second-order case: children above the age of *four* will usually say that Sally will look in the basket, which is where Sally (falsely) believes the marble to be. But younger children will usually say that Sally will look in the box: this is where the marble is, but Sally does not know this, and hence the answer is incorrect. Once again, to pass the test, the experimental subject must ascribe a *false* belief, this time to Sally. This ensures that the subject's answer can't be explained as the subject simply reporting what is true. For children with ASD, the shift to correct responses usually occurs at a later age.

Handling first-order false-belief tasks correctly is viewed as a milestone in the acquisition of *Theory of Mind (ToM)*, the ability to ascribe mental states such as beliefs to oneself and others, and some researchers account for ASD using what is called the *ToM deficit hypothesis* (see [2]). A wide range of first-order false-belief tasks have been devised, and over the past 30 years both correlational and training studies (involving both typically developing and children with ASD) have yielded robust results across various countries and various task manipulations; see, for example, the meta-analysis [22].

Second-order false-belief mastery, the topic of this paper, is also regarded as a key step in the acquisition of ToM, but much less is known about it, and many conclusions are tentative [13, 14]. There are far fewer second-order tests (the four we shall discuss pretty much cover the entire range) and they are less varied in design than their first-order cousins.[1] Moreover, there is no consensus on the status of the shift from first-order to second-order competency. Some researchers, starting with [20], have viewed it as a straightforward extension of first-order mastery: acquisition of second-order mastery occurs when the child has sufficiently strengthened his or her information processing capacities; following Miller [13, 14] we call this the *complexity only* position. Other researchers, starting with [15], have argued that the transition marks a more fundamental cognitive shift; following Miller, we call this the *conceptual change* position. In this paper we argue for a version of the conceptual change position. Our argument is grounded in ideas from modal and hybrid logic, but the backdrop to our discussion is our on-going training study on Danish speaking children with ASD in which we investigate whether training in linguistic recursion can lead to improvement in second-order false-belief competency.

We proceed as follows. In Section 2 we note earlier work on logical analysis of false-belief tasks. In Section 3 we give a coarse-grained modal-logical analysis of the four tasks and show that they share a common logical structure. A crucial role is played by a "principle of inertia", which says that an agent's belief is preserved over time, unless the agent gets information to the contrary. The coarse-grained analysis also reveals informational symmetries: all four possibilities inherent in the two binary dimensions of deception versus no-deception and change-in-world versus change-in-belief-only are realized. Moreover, the analysis reveals a somewhat puzzling feature concerning first-order information shared by all four tasks. In Section 4 we develop the coarse analysis into a fine-grained analysis by asking: how exactly is the principle of inertia used? Whereas the coarse-grained analysis simply uses the fact that Anne's belief is preserved, the fine-grained analysis builds on the observation that *Anne thinks that* Sally's belief is preserved to explain why; this is our stepping stone to the nested perspectival analyses that we formalize in hybrid-logic. In Section 5 we present the relevant fragment of hybrid logic, and formalize the first-order Sally-

[1] The three other task we consider are the bake-sale task, the ice-cream task, and the puppy task; see Tables 4, 5 and 6 in the Appendix. The ice-cream task was the very first second-order false-belief task to be used; it was introduced in 1985 by Wimmer and Perner in [15]. The bake-sale task is a variant of the ice-cream task, and, as is explained in [12], pages 323–324: "The stories were modeled after Wimmer and Perner's (1985) "ice cream truck story". In contrast to their stories, we made sure that the beliefs of the two main protagonists in the story did not overlap, both at first-order and second-order level: each protagonist had his or her own distinct belief which was different from that of the other protagonist, as well as from the belief of the participants." The puppy task was introduced in 1994 in [20], again as a simplification of Wimmer and Perner's ice-cream task.

Anne task. In Section 6 we extend this to a formalization of the second-order task; as we shall see, the proof modelling the first-order solution is nested inside the proof modelling the second-order solution. In Section 7 we conclude.

2 Logic and false-belief tasks

Frege and Husserl both tried to divorce logic and psychology, but post-1945 work in cognitive science and artificial intelligence put logic-based models of cognitive abilities back on the agenda, and the 2008 publication of Stenning and Van Lambalgen [19] brought logic and psychology even closer. This pioneering work considers a wide range of psychological tasks, including the first-order Sally-Anne tasks, which it analyses using non-monotonic closed-world reasoning. Stenning and Van Lambalgen make use of the principle of inertia, and draw a useful distinction between *belief formation* and *belief manipulation*, which we will adopt in our discussion below. The first-order Sally-Anne task has also been formalized using an interactive theorem prover for a many-sorted first-order modal logic, an approach which also makes use of the principle of inertia; see [1]. But we know of few examples of logical modelling of second-order false-belief tasks: the clearest is the Dynamic Epistemic Logic based analysis given in [4], though the use of game theory in [21] to investigate performance in higher-order social reasoning, for instance, is also relevant.

This paper builds on recent hybrid-logical work on false-beliefs [6, 7, 8]. The distinguishing feature of the hybrid-logical approach is that *perspective shift* is taken as fundamental. That is, it formalizes the local shifts of perspective required by the experimental subject when reasoning about the agents in the scenario (in our running example, Sally and Anne). The intuition is this: correctly handling the first-order Sally-Anne task seems to involve taking the perspective of Sally, and reasoning about what she believes. So to speak, you have to put yourself in Sally's shoes. As we shall argue below, correctly handling the second-order Sally-Anne task seems to involve taking the perspective of another agent, namely Anne, and reasoning about her perspective on Sally's belief: you have to put yourself in Anne's shoes while she is putting herself in Sally's shoes. In this paper we turn these shoes into nested natural deduction proofs.

3 A coarse-grained analysis

We now give a course-grained analysis of the four second-order tasks: we isolate the belief-states involved, and informally describe the reasoning leading from one belief-state to another. Three distinct times (t_0, t_1, and t_2) are significant in each story,[2] and in Table 2 in the Appendix we have described the belief-states at each of these times; the logical symbolism should be self-explanatory.

The reasoning pattern underlying all four tasks is clear.[3] First, note that in all four examples we make use of $B\neg\psi \rightarrow \neg B\psi$, the (contraposed form) of a modal principle called D: if we believe ψ to be false then we don't believe ψ. But in all four cases the crucial ingredient is the application of a "principle of inertia" saying that an agent's belief is preserved over time unless the agent has information to the contrary. For example, in the second-order Sally-Anne task, it is initially the case that Anne believes that Sally thinks that the marble is in the basket, formalized as $\boldsymbol{B_{anne}B_{sally}basket(t_0)}$. Initially it is also the case that Sally thinks that the marble is in the basket, $\boldsymbol{B_{sally}basket(t_0)}$, but Sally's belief changes at the intermediate stage t_1 since she sees through the keyhole the marble being moved, so $\boldsymbol{\neg B_{sally}basket(t_2)}$. Anne, however, does not know that Sally saw this, so Anne continues to believe that Sally thinks that the marble is in the basket, hence $\boldsymbol{B_{anne}B_{sally}basket(t_2)}$, the correct answer to the task.

This pattern underlies all four tasks: the correct answer is always a formula of the form $\boldsymbol{B_xB_y\phi}$ whose truth is preserved from stage t_0 to stage t_2, and subformula $\boldsymbol{B_y\phi}$ always becomes false at

[2] Some stories use more times than this: the bake-sale story, for example, makes use of (at least) four. But the sequence t_0, t_1, and t_2 constitutes the narrated time of the story, and here it is pointless to distinguish the times when Sam and Maria learn that there are no chocolate cookies for sale.

[3] In this section we adopt the following convention: belief-states that are part of this common pattern are typeset in bold (and displayed in blue in the online version), other belief-states are typeset in normal font.

stage t_1 — unbeknownst to agent x, who ends up in t_2 with a false belief about the belief of agent y. So to derive the correct answer $\boldsymbol{B_x B_y \phi(t_2)}$, the experimental subject must work out that agent x does not know that something led to a changed belief for agent y.

Zero-order, first-order and second-order information

Let's dig a little deeper for commonalities and differences. Table 3 in the Appendix summarizes the potentially relevant information available in the tasks, not just the information used in the coarse-grained analysis.

We start with the Sally-Anne task, where $\boldsymbol{B_x B_y \phi(t_2)}$ is instantiated to $\boldsymbol{B_{anne} B_{sally}\, basket(t_2)}$. Note that in Table 3 we have focused solely on the predicate occurring in the correct answer, namely *basket*, and ignored the predicate *box*. That is, we assume that what matters is whether or not Sally believes the marble has been moved from the basket, not where it has been moved to. With this restriction, rows 1–5 in Table 3 summarize the potentially relevant zero-order, first-order and second-order information in the Sally-Anne task.

Similarly, rows 6–10 in Table 3 summarize the information in the bake-sale task, where $\boldsymbol{B_x B_y \phi(t_2)}$ is instantiated to $\boldsymbol{B_{maria} B_{sam}\, chocolate(t_2)}$. We have again focussed on the predicate occurring in the correct answer, which in this case is *chocolate*, so we are assuming that what matters is whether or not chocolate cookies are for sale, not what else is. We have also restricted our attention to Maria and Sam, the agents involved in the correct answer, and ignored Mom and the mailman, as their perspectives seem irrelevant.

In a similar fashion, rows 11–15 and 16–20 summarize the information available in the ice-cream and the puppy tasks. Here we also restrict attention to the predicate ϕ and the agents x and y involved in the correct answer $\boldsymbol{B_x B_y \phi(t_2)}$. These restrictions enable us to compare the information in the various tasks in a uniform way, which we will now do.[4]

Let's start by comparing second-order information. First, note that in the Sally-Anne case, there is an asymmetry in the agents' second-order information (see rows 4 and 5): from time t_1 on, Sally believes that Anne believes that the marble has been moved away from the basket, since Sally can see Anne moving the marble. But Anne is not aware of this (Anne is deceived). On the other hand, in the bake-sale case, the second-order information (rows 9 and 10) is symmetric: at all three times Maria believes that Sam believes they sell chocolate cookies, and Sam also believes that Maria believes they sell chocolate cookies (so there is no deception).[5] So second-order information in the bake-sale case is *symmetric* whereas in the Sally-Anne case it is not. Similarly, the ice-cream task is symmetric (rows 14 and 15), but the puppy task is not (rows 19 and 20).

Next, let's consider the zero-order information. In the Sally-Anne case we have $basket(t_0)$, $\neg basket(t_1)$, and $\neg basket(t_2)$ (see row 1 of Table 3). So the formula $\boldsymbol{B_y \phi}$ becomes false at t_1 since both the world and the belief agent y has about the world change. On the other hand, in bake-sale we have $\neg chocolate(t_0)$, $\neg chocolate(t_1)$, and $\neg chocolate(t_2)$ (see row 6). In this case, the falsification of $\boldsymbol{B_y \phi}$ at t_1 is not caused by a change in the world, but only by a change in the belief agent y has about the world. We shall say that there is a *change-in-the-world* in the Sally-Anne case, and a *change-in-belief-only* in the bake-sale case. Similarly, there is a change-in-the-world in the ice-cream task (row 11), but a change-in-belief-only in the puppy task (row 16).

Table 3 sums up the zero-order and the second-order informational differences between the tasks. It shows that the bake-sale and (second-order) Sally-Anne tasks are maximally different — they differ both at zero-order and second-order levels — as are the ice-cream and the puppy stories.

[4] Note that rows 4,9,14,19 in Table 3 have the same form $\boldsymbol{B_x B_y \phi(t_0)}$, $\boldsymbol{B_x B_y \neg\phi(t_1)}$, $\boldsymbol{B_x B_y \neg\phi(t_2)}$ and are part of the common reasoning pattern leading to the correct answer, hence they are typeset in bold (and blue in the online version). Similarly, rows 2,7,12,17 have the same form $\boldsymbol{B_y \phi(t_0)}$, $\boldsymbol{B_y \neg\phi(t_1)}$, $\boldsymbol{B_y \neg\phi(t_2)}$ and are part of the common reasoning pattern, so they are also bold (and blue). That is, the information in these rows is part of the experimental design, and is intended to ensure that agent x ends up having a false belief about the belief of agent y.

[5] The distinction between tasks that do and do not involve deception is considered important for first-order false beliefs, as deception in a story may signal the relevance of detecting falsehood. But it has been little discussed for second-order tasks; see [14], especially pages 48-49, for discussion and pointers to the literature.

Table 1: Two dimensions of information variation

Task	Zero-order information	Second-order information
Ice-cream	Change-in-world	Symmetry
Bake-sale	Change-in-belief-only	Symmetry
Sally-Anne	Change-in-world	Asymmetry (deception)
Puppy	Change-in-belief-only	Asymmetry (deception)

Analyzing the first-order information reveals something curious. First, observe that in all four tasks we have $B_x \neg \phi(t_2)$. So at the last stage t_2 of each story, agent x believes — indeed *knows* — that ϕ is false. For example, in the Sally-Anne case, Anne knows that the marble is not in the basket (as she has moved it), and in the bake-sale case, Maria knows that no chocolate cookies are for sale. But in all four tasks we also have $B_x B_y \phi(t_2)$.

That is: in all four tasks there are false beliefs in *two* layers: there is the *outer* layer where the experimental subject has to ascribe a false belief to agent x, but there is also an *inner* layer where agent x ascribes a belief in a proposition to agent y, but agent x knows that this proposition is false. To put it another way: what we might call *inner first-order deception* is built into all four tasks. Note that this is different from the overt second-order deception present in the Sally-Anne and Puppy tasks: second-order deception plays a clear role in their experimental designs. But this inner first-order deception does *not* seem to be a part of the experimental design of the four second-order false belief tasks: $B_x \neg \phi(t_2)$ is *not* used to derive the correct answers.[6] Nonetheless, it seems hard to devise second-order scenarios which don't have inner first-order deception built into them without the experimental design falling apart. But as far as we are aware, the general presence of this kind of 'deception' is not something that has been noted or discussed in the literature on second-order false-beliefs.

4 A fine-grained analysis

We now make the coarse-grained analysis fine-grained by examining the role of the principle of inertia in more detail. Consider how it is used in the first-order Sally-Anne task. There the child (who we will call Peter) is asked: *Where will Sally look for her marble?* The inertia principle is clearly involved in Peter's reasoning: he takes it for granted that it can be applied to Sally's understanding. Indeed, learning to take it for granted in such circumstances is part of what is meant by acquiring first-order false-belief competence.

In the second-order case, Peter is asked: *Where does Anne think Sally will look for her marble?* Now, this is a question about Anne, thus it might seem that the key reasoning step for Peter is (once again) to take for granted that inertia applies, this time to *Anne's* understanding. After all, Anne never leaves the room, so she is right in front of that marble all the time, so inertia seems relevant. And as Peter observes, Anne does *not* "receive information to the contrary" (because she does not see Sally peek) and so the inertia principle applies and Anne's belief about Sally's belief will be preserved from t_0 to t_2.

But this analysis does not go deep enough. How does Peter "observe" that Anne does not "receive information to the contrary"? He certainly observes that Anne does not see Sally peek—*but what links this observation with Anne's beliefs?* There is a gap here. Peter cannot simply apply the principle of inertia to Anne's understanding; rather, he must understand that Anne is applying inertia to Sally's understanding. Anne reasons that Sally will preserve her belief in the marble being in the basket, *because Anne believes that Sally does not see the marble being moved.* This belief fills the missing gap—it builds a logical "bridge" to Peter's observation.

Summing up, in the fine-grained analysis the principle of inertia is applied *by Anne to Sally's belief* (and not by Peter to Anne's belief). And this has an interesting consequence. It means

[6] Which is why this information is not typeset in bold (and why in the online version, it is not in blue), and also why in Table 2 (the coarse-grained reasoning analysis) it has been put in parentheses.

that Anne is playing the same role in the second-order Sally-Anne task (namely, reasoning about Sally's belief) that Peter played in the first-order task. And this suggests a road to formalization: take a proof that formalizes the first-order task (Peter's reasoning about Sally) and view it instead as formalizing Anne's reasoning about Sally. Nest this proof (at the appropriate place) inside a formalization of Peter's reasoning about the second-order task; this will fill in the missing details about Anne's use of the inertia principle. This is the goal of the following two sections, where we will use natural deduction in hybrid logic to formalize the perspectival reasoning involved.

5 Formalizing the first-order Sally-Anne task

First we define the syntax and semantics of the fragment of hybrid logic we use for the formalization of the first-order Sally-Anne task, namely a version of Seligman's [18] *Logic of Correct Description (LCD)*. We assume we are given a set of propositional symbols (to be thought of as placeholders for information that is seen, believed, deduced . . . , and so on) and a set of nominals (to be thought of as names of the agents in the scenarios: Sally and Anne in our running example). We assume these sets are disjoint. We use p, q, r, . . . , for ordinary propositional symbols and s, a, b, c, . . . , for nominals.

Definition 5.1 *Formulas of LCD are defined by the following grammar:*

$$S \ ::= \ p \mid a \mid S \wedge S \mid S \rightarrow S \mid \bot \mid @_a S$$

Negation is defined by the convention that $\neg\phi$ is an abbreviation for $\phi \rightarrow \bot$.

Definition 5.2 *A* model *for LCD is a tuple* $(W, \{V_w\}_{w \in W})$ *where:*

1. *W is a non-empty set; think of these as the agents in the scenario of interest.*
2. *For each w, V_w is a function that to each ordinary propositional symbol assigns an element of* $\{0, 1\}$.

Given a model $\mathfrak{M} = (W, \{V_w\}_{w \in W})$, an *assignment* is a function g that to each nominal assigns an element of W. The relation $\mathfrak{M}, g, w \models \phi$, where g is an assignment, w is an element of W, and ϕ is a formula, is defined as follows:

$$\begin{array}{rll}
\mathfrak{M}, g, w \models p & \text{iff} & V_w(p) = 1 \\
\mathfrak{M}, g, w \models a & \text{iff} & w = g(a) \\
\mathfrak{M}, g, w \models \phi \wedge \psi & \text{iff} & \mathfrak{M}, g, w \models \phi \text{ and } \mathfrak{M}, g, w \models \psi \\
\mathfrak{M}, g, w \models \phi \rightarrow \psi & \text{iff} & \mathfrak{M}, g, w \models \phi \text{ implies } \mathfrak{M}, g, w \models \psi \\
\mathfrak{M}, g, w \models \bot & \text{iff} & \text{falsum} \\
\mathfrak{M}, g, w \models @_a\phi & \text{iff} & \mathfrak{M}, g, g(a) \models \phi
\end{array}$$

Two remarks. First, nominals should be thought of as naming the unique agent they are true at. For example, we shall use s as a nominal true at Sally; in effect it is a 'name' or 'constant' that picks her out.[7] But nominals are also used to make modalities: if ϕ is an arbitrary formula and s is the nominal that names Sally, then a new formula $@_s\phi$ can be built. The $@_s$ prefix is called a *satisfaction operator* and the formula $@_s\phi$ is called a *satisfaction statement.* Satisfaction statements let us switch perspectives: if we evaluate the satisfaction statement $@_s\phi$ at *any* agent in a model, it will be true iff ϕ is true at Sally.

Second, note that we have not introduced any modalities apart from the satisfaction operators. But this is not an oversight. In what follows the reader will encounter expressions of the form $@_sS\phi$ (that is, Sally sees ϕ) and $@_sB\phi$ (that is, Sally believes ϕ). But as far as the analysis of

[7] There are some interesting possibilities here: we could make our formalization more fine-grained by taking some nominals to stand for times, or go two-dimensional by taking nominals to stand for person-time pairs. But here we stick with the simpler setup just defined, as it has the same granularity as Stenning and Van Lambalgen's work on first-order false-belief tasks, cf. [19], pages 251–259.

Figure 1: Natural deduction rules for LCD

$$\frac{c \qquad \phi}{@_c\phi}\ (@I) \qquad\qquad \frac{c \qquad @_c\phi}{\phi}\ (@E)$$

$$\frac{\phi_1 \quad \dots \quad \phi_n \qquad\qquad \begin{array}{c}[\phi_1]\dots[\phi_n][c]\\ \vdots\\ \psi\end{array}}{\psi}\ (\mathit{Term})^* \qquad\qquad \frac{\begin{array}{c}[c]\\ \vdots\\ \psi\end{array}}{\psi}\ (\mathit{Name})^\dagger$$

$*$ $\phi_1, \dots, \phi_n$ and ψ are satisfaction statements, and there are no undischarged assumptions in the derivation of ψ besides the specified occurrences of $\phi_1, \dots, \phi_n$ and c.
† The nominal c does not occur in ψ or in any undischarged assumptions other than the specified occurrences of c.

first-order false-belief tasks is concerned, expressions containing the modalities S and B are not used in genuinely modal reasoning. Indeed, expressions of the form $S\phi$ and $B\phi$ are essentially complicated-looking propositional symbols: they are only used in simple propositional reasoning and then fed (once) into a perspective-shifting natural-deduction rule called *Term*. This will change (at least for the B operator) in the following section when we formalize the second-order task.

This brings us to natural deduction system we shall use to analyse the first-order Sally-Anne task.[8] We use the system for *LCD* obtained by extending the standard natural deduction system for classical propositional logic with the rules in Figure 1; the symbol c is an arbitrary nominal (that is, the name of an arbitrary agent). This is a modified version of Seligman's original natural deduction system for *LCD* [18]; these rules here are from Chapter 4 of [5]. We omit the rules for the boolean connectives: they are standard, and we prefer the more perspicuous proof trees obtained by 'compiling down' the simple propositional reasoning involved into additional rules (see the examples in the Appendix). In [5], this natural deduction system is proved to be sound and complete:

Theorem 5.3 *Let ψ be a formula and Γ a set of LCD wffs. The first statement below implies the second statement (soundness) and vice versa (completeness).*

1. *The formula ψ is derivable from Γ in Seligman's natural deduction system.*
2. *For any model $\mathfrak{M}$, any world w, and any assignment g, if, for any formula $\theta \in \Gamma$, it is the case that $\mathfrak{M}, g, w \models \theta$, then $\mathfrak{M}, g, w \models \psi$.*

Let's take a closer look. The rules $@I$ and $@E$ in Figure 1 are the introduction and elimination rules for satisfaction operators. The $@I$ rule says that if we have the information c (so we are reasoning about the agent called c) and we also have the information ϕ, then we can introduce the satisfaction operator $@_c$ and conclude $@_c\phi$, which says that ϕ holds from c's perspective. The $@E$ rule says: suppose that when reasoning about the agent named c, we also have the information that $@_c\phi$. Then we can eliminate $@_c$ and conclude ϕ.

But it is the *Term* rule that is central to the formalization. This rule lets us switch to another agent's perspective using hypothetical reasoning: the bracketed expressions $[\phi_1]\dots[\phi_n][c]$ in the statement of the rule are (discharged) assumptions. The key assumption is c, which can be glossed as: let's switch perspective and temporarily adopt c's point of view.[9] The remaining (discharged)

[8] Natural deduction was originally developed to model mathematical argumentation, but there is now some experimental backing for the claim that it is a mechanism underlying human deductive reasoning more generally; see [17]. One of the reasons we chose hybrid logic for our analysis (rather than, say, a multi-agent doxastic logic) was because of its well-behaved natural deduction systems; see [5].

[9] Incidentally, when using the *Term* rule we make at least one assumption c, but we can make several, and this is often necessary to drive the proof through.

assumptions $[\phi_1] \ldots [\phi_n]$ in the rule's statement are additional assumptions we may wish to make about the information available from c's perspective.[10]

The rule works as follows. Suppose that on the basis of assumptions $\phi_1 \ldots \phi_n, c$ we deduce ψ from c's perspective. Then the *Term* rule tells us that if $\phi_1 \ldots \phi_n$ are available in the original perspective,[11] then we can discharge the assumption (which we do by bracketing them, thus obtaining $[\phi_1] \ldots [\phi_n][c]$) and conclude ψ unconditionally in the original perspective.

The *Term* rule is a subtle and powerful rule.[12] Indeed, as was first shown in [6], the hybrid logical analysis of the first-order Sally-Anne task boils down to a *single* application of *Term*. Recall that Peter is the child performing the task. To answer the question (*Where will Sally look for her marble?*) Peter reasons as follows. At the time t_0, Sally believed the marble to be in the basket. She saw no action to move it, so she still believed this at t_1. When she returned at t_2, she still believed the marble to be in the basket (after all, she was out of the room when Anne moved it at time t_1). Peter concludes that Sally believes that the marble is still in the basket.

To formalize this we use the nominal s to name Sally, and the modal operators S (*sees that*) and B (*believes that*). The predicate $l(i,t)$ means that *the marble is at location i at time t*. Predicate $m(t)$ means that *the marble is moved at time t*. We take time to be discrete, and use $t+1$ as the successor of t. Using this vocabulary we can express the four belief formation principles we need:[13]

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

With the help of these principles, the perspectival reasoning involved in the Sally-Anne task can be formalized as the derivation in Figure 3 (in the Appendix). We have already given Peter's informal perspectival reasoning; the formal proof mirrors it in full detail using a single application of *Term* in which the assumptions of s model the shift to Sally's perspective. The first two premises $@_s Sl(\mathit{basket}, t_0)$ and $@_s S\neg m(t_0)$ taken together 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. The third premise, $@_s\neg Sm(t_1)$, says that Sally did not see the marble being moved at the time t_1 (since she was absent). Note that when applying the belief formation principles, we simply use them as rules.[14]

The bulk of the reasoning on the right-hand-side of the proof tree in Figure 3 simply consists of a sequence of applications of belief formation principles until the crucial formula $@_s Bl(\mathit{basket}, t_2)$ — *Sally believes the ball is in the basket* — is deduced. What turns this into a formalisation of correct reasoning in the Sally-Anne task is the way the sequencing of belief formation principles is perspectivized. The right-hand-side sequencing occurs between the initial assumptions of s (which perspectivizes it as *Sally's* reasoning) and the final application of *Term* which lets us conclude

[10] The *Name* rule tells us that if we can prove the information ϕ by adopting some *arbitrary* perspective c, then ϕ also holds from the original perspective. As we won't use this rule in our analysis, we refer to [5] for further discussion.

[11] Indicated by the premisses $\phi_1 \ldots \phi_n$ listed just above the horizontal line in the statement of *Term* given in Figure 1.

[12] A subtlety worth emphasising is that (as is stated in Figure 1) the assumptions $[\phi_1] \ldots [\phi_n]$ must all be satisfaction statements, otherwise the rule is not sound. We refer the reader to [18] and Chapter 4 of [5] for further discussion.

[13] As we mentioned earlier, "belief formation" (and "belief manipulation") is terminology we have borrowed from [19], and we discuss them in more detail shortly. As for the belief formations principles themselves, we have already met Principle (D) which says if we believe that something is false, then we don't believe it. Principle $(P1)$ states that a belief in ϕ may be formed as a result of seeing ϕ; this is principle (9.2) in [19], page 251. Principle $(P2)$ is (pretty clearly) a principle of inertia: a belief that the predicate l is true is preserved from a time t to its successor $t+1$, unless it is believed that the marble moved at t. This is essentially Principle (9.11) from [19], page 253, and axiom $[A_5]$ in [1], page 20. 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. Obviously this is not a general truth, but the point of the formalization is simply to capture Peter's reasoning in the Sally-Anne scenario.

[14] As we remarked earlier, we do this to 'compile down' the simple propositional reasoning involved. Strictly speaking, deducing $B\phi$ from $S\phi$ requires us to apply the propositional rule of *modus ponens* to $S\phi \rightarrow B\phi$. Using the belief formation principles as additional natural deduction rules enables us to omit such steps and reduce the size of the proof tree.

Figure 2: Belief manipulation rule for the B operator

$$\dfrac{B\phi_1 \quad \dots \quad B\phi_n \qquad \begin{array}{c}[\phi_1]\dots[\phi_n]\\ \vdots \\ \psi\end{array}}{B\psi}\ (\mathsf{BM})^*$$

$*$ There are no undischarged assumptions in the derivation of ψ except the specified occurrences of $\phi_1, \dots, \phi_n$.

that the crucial formula is also true from *Peter's* point of view. In short, the analysis consists of *Belief Formation + Perspectival Reasoning* correctly combined.

Analagous remarks are made by Stenning and Van Lambalgen about their own analysis of first-order false-belief tasks; see [19], page 257. They note that the bulk of the reasoning involves belief formation principles and their analysis succeeds because it is carrying out using closed world reasoning; we might summarise their approach as *Belief Formation + Closed World Reasoning* correctly combined. However they then go on to remark that what they call *Belief Manipulation* rules (which codify how to reason from one belief state to another) are unnecessary. Now, as far as first-order false-belief reasoning is concerned, we agree completely. Indeed, until now we have provided no proof rules for manipulating the belief operator B beyond the belief formation principles. And that is because, for the first-order Sally-Anne task, we had no need of anything else. But a belief manipulation rule will be needed if we are to extend our perspectival analysis to the second-order Sally-Anne task.[15] We turn to this task now.

6 Formalizing the second-order Sally-Anne task

As we remarked at the end of Section 4, Anne plays the same role in the second-order Sally-Anne task (namely, reasoning about Sally's belief) that Peter played in the first-order task. This suggests that we should take the proof we have just given (formalizing Peter's reasoning about Sally), view it as formalizing Anne's reasoning about Sally, and nest it (at the appropriate place) inside a formalization of Peter's reasoning about the second-order task. That is, we should add another level of nesting to the perspectival analysis. To make this work we have to introduce a recursive belief manipulation rule for B. We have chosen the rule given in Figure 2. We call it BM. It is a version of a rule from [10] that fits naturally our tree-style natural deduction proofs.[16]

And now to complete the formalization. We shall use the nominal a as a name for Anne, read $D\phi$ as ϕ is deducible, and make use of a natural deduction formulation of the following belief formation principle:

$(P0) \qquad D\phi \rightarrow B\phi.$

This says that if we can deduce the information ϕ then we believe ϕ (this is principle (9.4) in [19], page 251). With this machinery in place, the reasoning in the second-order Sally-Anne task can be formalized by the proof tree in Figure 4 in the Appendix. Note that the first-order proof in nested inside: the dots in the upper-right corner of Figure 4 indicate where.

The proof's conclusion, $@_a B @_s Bl(\mathit{basket}, t_2)$, says that Anne believes that Sally believes that the marble is in the basket at the time t_2, and this is indeed the correct response to the second-order task. And Peter can prove this as follows.

The first two premises used in the application of *Term* with which the proof concludes, $@_a S @_s Sl(\mathit{basket}, t_0)$ and $@_a S @_s S\neg m(t_0)$, say that at time t_0, Anne saw that Sally saw that the marble was in the basket and that no action was taken to move it. The third premise used in the

[15] Stenning and Van Lambalgen do not analyse second-order false-belief tasks.

[16] So we are adding natural deduction machinery for the minimal modal logic K and thus treating B as a full-fledged modal operator. In this paper we won't discuss the model-theoretic changes required — but we *do* believe that the fact that a semantic enrichment is called for at this point adds weight to our argument that the transition from first- to second-order reasoning involves conceptual change.

concluding application of the *Term* rule, $@_a D @_s \neg Sm(t_1)$, says that Anne deduced that Sally did not see the marble being moved at the time t_1, which is true.

But the essential step is the way the belief manipulation rule BM glues together the two levels of perspectival reasoning. The embedded proof (which reasons from Sally's perspective) yields the conclusion $@_s Bl(basket, t_2)$, the correct response to the first-order task. But Peter can't use this information directly: he needs to know that *Anne believes this.* But the application of BM prefixes the belief operator to form $B@_s Bl(basket, t_2)$, and the very next step of the proof shows that this belief holds from Anne's point of view. Thus the reasoning on the right has now been incorporated back into Anne's perspective, and so can be fed into *Term*, and Peter has his answer.

7 Concluding discussion

Second-order reasoning is more complex than first-order — the previous section with its embedded proof and use of the BM rule showed this clearly.[17] Nonetheless, our analysis also suggests that the transition to second-order competence marks a more significant development than is suggested by the complexity only position: the full *reification* of beliefs. Attainment of first-order false-belief competence marks the stage at which the child becomes aware of the fact that beliefs held by other agents can be false; second-order competence, on the other hand, marks the stage where beliefs become objects in their own right that can be manipulated. This shift is mirrored in our analysis: we jumped from a logic that permitted only *Belief Formation + Perspectival Reasoning* to one that allowed unrestricted *Belief Manipulation* as well.

This is a significant advance. Beliefs are special objects: they are abstract, invisible, and though 'about' the world, they may very well be false. Typically developing children learn this first lesson around the age of four, but there is a further lesson they must learn: that beliefs can be embedded one inside another and freely manipulated. Something like the BM rule seems to be required to capture this step. It is tempting to speculate that at this developmental stage some sort of "recursion module" is adapted to handle these strange new objects, but be that as it may, in typically developing children the reasoning architecture is certainly enriched in an important way at around the age of six.[18]

Recursively stacked beliefs lie at the heart of this transition, which brings us to our empirical work [16]. Our logical investigations were carried out as part of an ongoing training study involving Danish speaking children with ASD. Our empirical work is driven by the hypothesis that, in case of children with ASD, improving linguistic recursion competency predicts belief manipulation mastery required by second-order false-belief tasks. We are investigating whether children with ASD use language as a "scaffolding" to support developing understanding of other minds, an explanation advanced in the first-order case by [11].

Acknowledgements

We thank the referees for their valuable comments and questions. The authors acknowledge the funding received from the VELUX FOUNDATION for the project *Hybrid-Logical Proofs at Work in Cognitive Psychology* (VELUX 33305).

[17] Indeed, our analysis allows us to tentatively indicate the shift in complexity. The *LCD* fragment is NP-complete. By adding BM we have moved to a PSPACE-hard modal logic. So our analysis of the first-order Sally-Anne task is carried out in computationally simpler logic than the second-order case (assuming P$\neq$ NP).

[18] Our formalization does suggest a hypothesis which may be empirically testable. Although we have talked of acquiring *second-order* competency, to acquire (something like) the BM rule is to acquire a fully recursive competency. That is, once the child has acquired BM, there should be nothing more to learn, for the rule covers the third, fourth, fifth, ..., and all higher-order levels. That is, we suspect that false-belief competency comes in two stages for typically developing children: first-order competency (at around the age of four) and all the rest (at around the age of six). But designing an experiment to test this is likely to be difficult. Apart from anything else, higher levels of reasoning impose heavy cognitive loads very fast, and it is unclear how such performance effects could be disentangled experimentally.

References

[1] K. Arkoudas and S. Bringsjord. Toward formalizing common-sense psychology: An analysis of the false-belief task. In *PRICAI 2008: Trends in Artificial Intelligence*, volume 5351 of *Lecture Notes in Computer Science*, pages 17–29. Springer-Verlag, 2008.

[2] Simon Baron-Cohen, Alan M Leslie, and Uta Frith. Does the autistic child have a 'theory of mind'? *Cognition*, 21(1):37–46, 1985.

[3] Simon Baron-Cohen, Michelle O'Riordan, Valerie Stone, Rosie Jones, and Kate Plaisted. Recognition of faux pas by normally developing children and children with Asperger syndrome or high-functioning autism. *Journal of Autism and Developmental Disorders*, 29(5):407–418, 1999.

[4] 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.

[5] T. Braüner. *Hybrid Logic and its Proof-Theory*, volume 37 of *Applied Logic Series*. Springer, 2011.

[6] T. Braüner. Hybrid-logical reasoning in the Smarties and Sally-Anne tasks. *Journal of Logic, Language and Information*, 23:415–439, 2014.

[7] 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 Meeting of the Cognitive Science Society*, pages 273–278. Pasadena, California: Cognitive Science Society, 2015.

[8] T. Braüner, P. Blackburn, and I. Polyanskaya. Recursive belief manipulation and second-order false-beliefs. In *Proceedings of the 38th Annual Meeting of the Cognitive Science Society*, pages 2579–2584. Philadelphia, Pennsylvania, USA: Cognitive Science Society, 2016.

[9] T. Braüner, P. Blackburn, and I. Polyanskaya. Second-order false-belief tasks: Analysis and formalization. In *Proceedings of Workshop on Logic, Language, Information and Computation (WoLLIC 2016)*, volume 9803 of *Lecture Notes in Computer Science*, pages 125–144. Springer-Verlag, 2016.

[10] M. Fitting. Modal proof theory. In P. Blackburn, J. van Benthem, and F. Wolter, editors, *Handbook of Modal Logic*, pages 85–138. Elsevier, 2007.

[11] C.H. Hale and H. Tager-Flusberg. The influence of language on theory of mind: a training study. *Developmental Science*, 6:346–359, 2003.

[12] B. Hollebrandse, A. van Hout, and P. Hendriks. Children's first and second-order false-belief reasoning in a verbal and a low-verbal task. *Synthese*, 191:321–333, 2014.

[13] S.A. Miller. Children's understanding of second-order mental states. *Psychological Bulletin*, 135:749–773, 2009.

[14] Scott Miller. *Theory of mind: Beyond the preschool years*. Psychology Press, 2012.

[15] 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.

[16] Irina Polyanskaya, Torben Braüner, and Patrick Blackburn. Linguistic recursion and Autism Spectrum Disorder. Manuscript, 2016.

[17] L.J. Rips. Logical approaches to human deductive reasoning. In J.E. Adler and L.J. Rips, editors, *Reasoning: Studies of Human Inference and Its Foundations*, pages 187–205. Cambridge University Press, 2008.

[18] J. Seligman. The logic of correct description. In M. de Rijke, editor, *Advances in Intensional Logic*, volume 7 of *Applied Logic Series*, pages 107 – 135. Kluwer, 1997.

[19] K. Stenning and M. van Lambalgen. *Human Reasoning and Cognitive Science.* MIT Press, 2008.

[20] K. Sullivan, D. Zaitchik, and H. Tager-Flusberg. Preschoolers can attribute second-order beliefs. *Developmental Psychology*, 30:395–402, 1994.

[21] J. Szymanik, B. Meijering, and R. Verbrugge. Using intrinsic complexity of turn-taking games to predict participants' reaction times. In M. Knauff, M. Pauen, N. Sebanz, and I. Wachsmuth, editors, *Proceedings of the 35th Annual Conference of the Cognitive Science Society*, pages 1426–1432. Cognitive Science Society, 2013.

[22] 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.

Appendix

The Appendix contains the coarse-grained reasoning for the four tasks (Table 2), the table listing their information content (Table 3), the texts of the bake-sale, ice-cream and puppy tasks (Tables 4, 5 and 6 respectively) and the formalization of the first-order and second-order Sally-Anne tasks (Figures 3 and 4).

Table 2: A coarse-grained analysis of second-order false-belief tasks in terms of belief-states

	Time t_0	*Time t_1*	*Time t_2*
Second-order Sally-Anne task	*Sally leaves after having put the marble in the basket* Anne now believes that Sally thinks that the marble is in the basket $\boldsymbol{B_{anne}B_{sally}basket(t_0)}$	*Anne moves the marble from the basket to the box* Sally sees through the keyhole that the marble is moved $\boldsymbol{B_{sally}box(t_1)}$ So $\boldsymbol{B_{sally}\neg basket(t_1)}$ and hence $\boldsymbol{\neg B_{sally}basket(t_1)}$ (Anne knows that the marble is moved so $B_{anne}\neg basket(t_1)$)	*Sally has returned* Correct answer: “Anne believes that Sally thinks that the marble is in the basket” $\boldsymbol{B_{anne}B_{sally}basket(t_2)}$ Derivable by inertia from t_0 as Anne does not know that Sally’s belief changed at t_1
Bake-sale task	*Maria tells Sam that she will go to buy chocholate cookies* Maria now believes that Sam thinks that they sell chocolate cookies $\boldsymbol{B_{maria}B_{sam}chocolate(t_0)}$	*Mom comes home and Maria arrives at the bake sale* Sam is told that they sell pumpkin pie $\boldsymbol{B_{sam}pumpkin(t_1)}$ So $\boldsymbol{B_{sam}\neg chocolate(t_1)}$ and hence $\boldsymbol{\neg B_{sam}chocolate(t_1)}$ (Maria realizes that they sell brownies so $B_{maria}\neg chocolate(t_1)$)	*The mailman talks to Maria* Correct answer to the mailman: “I [Maria] believe that Sam thinks that they sell chocolate cookies” $\boldsymbol{B_{maria}B_{sam}chocolate(t_2)}$ Derivable by inertia from t_0 as Maria does not know that Sam’s belief changed at t_1
Ice-cream task	*Mary leaves the park after having seen the van* John believes that Mary thinks that the van is in the park $\boldsymbol{B_{john}B_{mary}park(t_0)}$	*John and Mary independently talk to the ice-cream man* Mary is told that the van drives to the church $\boldsymbol{B_{mary}church(t_1)}$ So $\boldsymbol{B_{mary}\neg park(t_1)}$ and hence $\boldsymbol{\neg B_{mary}park(t_1)}$ (John is told that the van drives to the church so $B_{john}\neg park(t_1)$)	*The van has arrived* Correct answer: “John believes that Mary thinks that the van is in the park” $\boldsymbol{B_{john}B_{mary}park(t_2)}$ Derivable by inertia from t_0 as John does not know that Mary’s belief changed at t_1
Puppy task	*Mom tells Peter that she has got him a toy* Mom now believes that Peter thinks that he will get a toy $\boldsymbol{B_{mom}B_{peter}toy(t_0)}$ (Mom knows that Peter will get a puppy so $B_{mom}\neg toy(t_0)$ is true)	*Peter finds a puppy in the basement* Peter realizes that he will get a puppy $\boldsymbol{B_{peter}puppy(t_1)}$ So $\boldsymbol{B_{peter}\neg toy(t_1)}$ and hence $\boldsymbol{\neg B_{peter}toy(t_1)}$	*Grandmother talks to Mom* Correct answer to Grandmother: “I [Mom] believe that Peter thinks that he will get a toy” $\boldsymbol{B_{mom}B_{peter}toy(t_2)}$ Derivably by inertia from t_0 as Mom does not know that Peter’s belief changed at t_1

Table 3: Zero-order, first-order and second-order information in the tasks

Sally-Anne	*Time* t_0	*Time* t_1	*Time* t_2	
Zero-order	$basket(t_0)$	$\neg basket(t_1)$	$\neg basket(t_2)$	1
First-order	$\boldsymbol{B_{sally}\,basket(t_0)}$	$\boldsymbol{B_{sally}\neg basket(t_1)}$	$\boldsymbol{B_{sally}\neg basket(t_2)}$	2
	$B_{anne}\,basket(t_0)$	$B_{anne}\neg basket(t_1)$	$B_{anne}\neg basket(t_2)$	3
Second-order	$\boldsymbol{B_{anne}B_{sally}\,basket(t_0)}$	$\boldsymbol{B_{anne}B_{sally}\,basket(t_1)}$	$\boldsymbol{B_{anne}B_{sally}\,basket(t_2)}$	4
	$B_{sally}B_{anne}\,basket(t_0)$	$B_{sally}B_{anne}\neg basket(t_1)$	$B_{sally}B_{anne}\neg basket(t_2)$	5
Bake-sale				
Zero-order	$\neg chocolate(t_0)$	$\neg chocolate(t_1)$	$\neg chocolate(t_2)$	6
First-order	$\boldsymbol{B_{sam}\,chocolate(t_0)}$	$\boldsymbol{B_{sam}\neg chocolate(t_1)}$	$\boldsymbol{B_{sam}\neg chocolate(t_2)}$	7
	$B_{maria}\,chocolate(t_0)$	$B_{maria}\neg chocolate(t_1)$	$B_{maria}\neg chocolate(t_2)$	8
Second-order	$\boldsymbol{B_{maria}B_{sam}\,chocolate(t_0)}$	$\boldsymbol{B_{maria}B_{sam}\,chocolate(t_1)}$	$\boldsymbol{B_{maria}B_{sam}\,chocolate(t_2)}$	9
	$B_{sam}B_{maria}\,chocolate(t_0)$	$B_{sam}B_{maria}\,chocolate(t_1)$	$B_{sam}B_{maria}\,chocolate(t_2)$	10
Ice-cream				
Zero-order	$park(t_0)$	$\neg park(t_1)$	$\neg park(t_2)$	11
First-order	$\boldsymbol{B_{mary}\,park(t_0)}$	$\boldsymbol{B_{mary}\neg park(t_1)}$	$\boldsymbol{B_{mary}\neg park(t_2)}$	12
	$B_{john}\,park(t_0)$	$B_{john}\neg park(t_1)$	$B_{john}\neg park(t_2)$	13
Second-order	$\boldsymbol{B_{john}B_{mary}\,park(t_0)}$	$\boldsymbol{B_{john}B_{mary}\,park(t_1)}$	$\boldsymbol{B_{john}B_{mary}\,park(t_2)}$	14
	$B_{mary}B_{john}\,park(t_0)$	$B_{mary}B_{john}\,park(t_1)$	$B_{mary}B_{john}\,park(t_2)$	15
Puppy				
Zero-order	$\neg toy(t_0)$	$\neg toy(t_1)$	$\neg toy(t_2)$	16
First-order	$\boldsymbol{B_{peter}\,toy(t_0)}$	$\boldsymbol{B_{peter}\neg toy(t_1)}$	$\boldsymbol{B_{peter}\neg toy(t_2)}$	17
	$B_{mom}\neg toy(t_0)$	$B_{mom}\neg toy(t_1)$	$B_{mom}\neg toy(t_2)$	18
Second-order	$\boldsymbol{B_{mom}B_{peter}\,toy(t_0)}$	$\boldsymbol{B_{mom}B_{peter}\,toy(t_1)}$	$\boldsymbol{B_{mom}B_{peter}\,toy(t_2)}$	19
	$B_{peter}B_{mom}\,toy(t_0)$	$B_{peter}B_{mom}\neg toy(t_1)$	$B_{peter}B_{mom}\neg toy(t_2)$	20

Table 4: The bake-sale task (quoted from [12], pictures and some questions omitted)

Sam and Maria are playing together. They look outside and see that the church is having a bake sale. Maria tells Sam: "I am going to buy chocolate chip cookies for us there," and she walks away.

Mom comes home and she tells Sam that she just drove past the bake sale. "Are they selling chocolate chip cookies?" Sam asks. No, mum says, "they are only selling pumpkin pie." "Maria will now probably get pumpkin pie at the bake sale," Sam says.

Maria has arrived at the bake sale. "I would like to buy chocolate chip cookies," she says. "All we have left are brownies," says the lady behind the stall. Since Maria also likes brownies, she decides to get some brownies.

On her way back, Maria meets the mailman. She tells the mailman: "I have just bought some brownies. I am going to share them with my brother Sam. It is a surprise". "That is nice of you," says the mailman. Then he asks Maria: "Does Sam know what you bought him?"

Ignorance: What does Maria tell the mailman?

Then the mailman asks: "What does Sam think they are selling at the bake sale?"

Second-order false-belief question: What does Maria tell the mailman?

Table 5: The ice-cream task (introduced in [15], quoted from [13], a question omitted)

This is a story about John and Mary who live in this village. This morning John and Mary are together in the park. In the park there is also an ice-cream man in his van.

Mary would like to buy an ice cream but she has left her money at home. So she is very sad. “Don’t be sad,” says the ice-cream man, “you can fetch your money and buy some ice cream later. I’ll be here in the park all afternoon.” “Oh good,” says Mary, “I’ll be back in the afternoon to buy some ice cream. I’ll make sure I won’t forget my money then.”

So Mary goes home. . . . She lives in this house. She goes inside the house. Now John is on his own in the park. To his surprise he sees the ice-cream man leaving the park in his van. “Where are you going?” asks John. The ice-cream man says, “I’m going to drive my van to the church. There is no one in the park to buy ice cream; so perhaps I can sell some outside the church.”

The ice-cream man drives over to the church. On his way he passes Mary’s house. Mary is looking out of the window and spots the van. “Where are you going?” she asks. “I’m going to the church. I’ll be able to sell more ice cream there,” answers the man. “It’s a good thing I saw you,” says Mary. Now John doesn’t know that Mary talked to the ice-cream man. He doesn’t know that!

Now John has to go home. After lunch he is doing his homework. He can’t do one of the tasks. So he goes over to Mary’s house to ask for help. Mary’s mother answers the door. “Is Mary in?” asks John. “Oh,” says Mary’s mother. “She’s just left. She said she was going to get an ice cream.”

Test question: So John runs to look for Mary. Where does he think she has gone?

Table 6: The puppy task (introduced in [20], quoted from [13], some questions omitted)

Tonight it’s Peter’s birthday and Mom is surprising him with a puppy. She has hidden the puppy in the basement. Peter says, “Mom, I really hope you get me a puppy for my birthday.” Remember, Mom wants to surprise Peter with a puppy. So, instead of telling Peter she got him a puppy, Mom says, “Sorry Peter, I did not get you a puppy for your birthday. I got you a really great toy instead.”

Now, Peter says to Mom, “I’m going outside to play.” On his way outside, Peter goes down to the basement to fetch his roller skates. In the basement, Peter finds the birthday puppy! Peter says to himself, “Wow, Mom didn’t get me a toy, she really got me a puppy for my birthday.” Mom does not see Peter go down to the basement and find the birthday puppy.

Now, the telephone rings, ding-a-ling! Peter’s grandmother calls to find out what time the birthday party is. Grandma asks Mom on the phone, “Does Peter know what you really got him for his birthday?”

Now remember, Mom does not know that Peter saw what she got him for his birthday. Then, Grandma says to Mom, “What does Peter think you got him for his birthday?”

Second-order false-belief question: What does Mom say to Grandma?

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

$$\dfrac{@_s Sl(basket,t_0) \quad @_s S\neg m(t_0) \quad @_s \neg Sm(t_1) \qquad \dfrac{[s] \qquad \dfrac{\dfrac{\dfrac{\dfrac{[s]\ [@_s Sl(basket,t_0)]}{Sl(basket,t_0)}\,(@E)}{Bl(basket,t_0)}\,(P1) \qquad \dfrac{\dfrac{\dfrac{[s]\ [@_s S\neg m(t_0)]}{S\neg m(t_0)}\,(@E)}{B\neg m(t_0)}\,(P1)}{\neg Bm(t_0)}\,(\mathsf{D})}{Bl(basket,t_1)}\,(P2) \qquad \dfrac{\dfrac{[s]\ [@_s \neg Sm(t_1)]}{\neg Sm(t_1)}\,(@E)}{\neg Bm(t_1)}\,(P3)}{Bl(basket,t_2)}\,(P2)}{@_s Bl(basket,t_2)}\,(@I)}{@_s Bl(basket,t_2)}\,(\mathit{Term})$$

Figure 4: Formalization of the child's correct response in the second-order Sally-Anne task

$$\dfrac{@_a S @_s Sl(basket,t_0) \quad @_a S @_s S\neg m(t_0) \quad @_a D @_s \neg Sm(t_1) \qquad \dfrac{\dfrac{\dfrac{\dfrac{[a][@_a S @_s Sl(basket,t_0)]}{S @_s Sl(basket,t_0)}}{B @_s Sl(basket,t_0)}\,(P1) \quad \dfrac{\dfrac{[a][@_a S @_s S\neg m(t_0)]}{S @_s S\neg m(t_0)}}{B @_s S\neg m(t_0)}\,(P1) \quad \dfrac{\dfrac{[a][@_a D @_s \neg Sm(t_1)]}{D @_s \neg Sm(t_1)}}{B @_s \neg Sm(t_1)}\,(P0) \quad \begin{array}{c}[@_s Sl(basket,t_0)][@_s S\neg m(t_0)][@_s \neg Sm(t_1)]\\ \vdots \\ @_s Bl(basket,t_2)\end{array}}{B @_s Bl(basket,t_2)}\,(\mathsf{BM}) \qquad [a]}{@_a B @_s Bl(basket,t_2)}}{@_a B @_s Bl(basket,t_2)}\,(\mathit{Term})$$

The vertical dots in the upper-right corner represent the derivation in Figure 3. So this proof contains two applications of *Term*: the concluding application, which is shown, and the one inside the earlier proof, which is not. To save space, we have omitted names of the introduction and elimination rules for the @ operator.