



Helping as an early indicator of a theory of mind: Mentalism or Teleology?



Beate Priewasser^{a,b,1,*}, Eva Rafetseder^{c,1}, Carina Gargitter^b, Josef Perner^{a,b}

^a Centre for Cognitive Neuroscience, University of Salzburg, Austria

^b Department of Psychology, University of Salzburg, Austria

^c Faculty of Natural Sciences, University of Stirling, Scotland, United Kingdom

ARTICLE INFO

Keywords:

Helping paradigm

Theory of mind

Replication

Teleology

Early false belief understanding

ABSTRACT

This article challenges the claim that young children's helping responses in [Buttelmann, Carpenter, and Tomasello's \(2009\)](#) task are based on ascribing a false belief to a mistaken agent. In our first Study 18- to 32-month old children ($N = 28$) were more likely to help find a toy in the false belief than in the true belief condition. In Study 2, with 54 children of the same age, we assessed the authors' mentalist interpretation of this result against an alternative teleological interpretation that does not make the assumption of belief ascription. The data speak in favor of our alternative. Children's social competency is based more on inferences about what is likely to happen in a particular situation and on objective reasons for action than on inferences about agents' mental states. We also discuss the need for testing serious alternative interpretations of claims about early belief understanding.

1. Introduction

Buttelmann's helping paradigm plays an important role in the discussion about when infants or children come to understand belief ([Buttelmann, Carpenter, & Tomasello, 2009](#); BCT). In the standard false belief test children are asked to predict where an agent, who is mistaken about an object's location, will look for it. Quite reliably only by about 4 years children answer this question correctly ([Wimmer & Perner, 1983](#); [Wellman, Cross, & Watson, 2001](#)). In contrast, children's looking behavior that indicates their anticipations about where the agent will search for the object provides evidence for sensitivity to false beliefs in infants as young as 18 months ([Clements & Perner, 1994](#); [Southgate, Senju, & Csibra, 2007](#); [Thoermer, Sodian, Vuori, Perst, & Kristen, 2012](#)). In violation of expectation paradigms evidence was found at an even younger age around 14–16 months ([Onishi & Baillargeon, 2005](#); [Surian, Caldi, & Sperber, 2007](#)). Prolonged looking when an agent's belief does not match the child's own belief showed sensitivity to the agent's belief as young as 7 months and similar ages are reported for neural signatures of representing belief ([Kampis, Parise, Csibra, & Kovacs, 2015](#); [Kovacs, Teglas, & Endress, 2010](#); [Southgate & Vernetti, 2014](#)). BCT provide an importantly different kind of evidence for early understanding of belief because they used helping behavior, an intentional action, an indicator of understanding. Before expanding on the ongoing debate about the nature of young children's false belief understanding and on our alternative interpretation of BCT's findings we start by describing their procedure and interpretation in detail.

Two experimenters E1 and E2 (E2 being the *agent* to be helped) engage with the child C. After a short warm up E2 discovers two boxes of different color (A and B) and opens and closes the lids of both with interest. In E2's absence E1 shows C how the boxes can be

* Corresponding author.

E-mail address: beate.priewasser@sbg.ac.at (B. Priewasser).

¹ These authors contributed equally to this work.

locked and opened with a pin. E1 leaves the boxes unlocked and E2 returns excitedly with a caterpillar toy. She plays for a while with it, introducing it also to E1 and C, and puts it into box A. In the *false belief (FB) condition* E2 leaves again to get her keys. E1 and C “play a trick” on E2 and E1 sneakily moves the toy to box B, continuously checking on the door, and then locks both boxes with the pin. E2 returns and tries to open box A (where she thinks the toy still is). In the *true belief (TB) condition* E2 stays in the room and watches how E1 moves the toy to box B with mutual eye contact with E2 and C. E2 looks briefly away when E1 locks the boxes with the pin. After going to check whether the door was properly shut, E2 approaches box A and unsuccessfully tries to open it. If the child does not respond immediately, E1 suggests to C to help E2 and if needed, several prompts follow. BCT tested one group of 18-month- and another group of 30–32-month-old children and found that in each age group children in the FB condition tended to go to box B to retrieve the toy for E2, while in the TB condition they tended to help E2 to open box A. The mentalistic interpretation preferred by BCT is as follows: When E2 tries to open the locked box A the child has to figure out the reason for E2’s action. Since in the FB condition E2 thinks that the toy is still in box A she most likely is looking for her toy. Since she does not know where the toy really is the children help her find it in box B. In the TB condition E2 knows that the toy is in box B, therefore she cannot be after the toy when trying to open box A. She must be trying to open A for some other (unknown) reason.

BCT’s results play a central role for theories about the cognitive basis of early theory of mind competences. Two lines of explanation are particularly prominent. The first one distinguishes between an implicit and explicit understanding (Clements & Perner, 2001; Onishi & Baillargeon, 2005). Children’s looking is an indirect measure of their knowledge. They look as a consequence of their expectations and not in order to serve a purpose (They look there *because* the agent will go there, not *in order* that the agent will go there, nor *in order* to tell the experimenter what they are thinking). Whereas, they answer the test question posed in the traditional false belief task *in order* to answer the question. Appropriate responses on an indirect measure (looking) in the absence of correct responding to a direct measure (answer to question) is taken as a sign of implicit knowledge in the consciousness literature (Reingold & Merikle, 1993; applied to false belief studies: Clements & Perner, 2001). Children’s helping behavior in BCT’s study is highly relevant evidence, since it is a direct measure: children do it *in order* to help the experimenter. So their data speak against the implicit knowledge explanation (Carruthers, 2013, p. 145).

The second main explanation of the discrepancy in when children take belief into account assumes that children have explicit knowledge from early on but cannot show it in the standard task due to processing limitations. Baillargeon, Scott, and He (2010) distinguish between spontaneous and elicited test responses. Looking time and gaze direction are spontaneous responses, while the answers to the traditional test questions are elicited responses. In particular the additional processing required by the test question is supposed to exceed younger children’s processing capacity. BCT’s finding thus poses a problem for this theory since children’s helping is elicited by E1’s verbal suggestion to help E2 (the relevant agent to whom a belief is supposed to be attributed). Hence it should be as difficult as the standard test, which it is not. To account for BCT’s data Carruthers (2013, p. 152), for instance, saw the need to amend Baillargeon et al.’s theory with assumptions from language pragmatics as proposed by Helming, Strickland, and Jacob (2014) and Helming, Strickland, and Jacob (2016). When being asked a question by the experimenter in the traditional test, children have to coordinate their third-person perspective as a listener to the story with their second person perspective when interacting with the experimenter. It is this coordination of perspectives that makes the traditional task so difficult and helping in BCT’s procedure easy since this does not necessitate such coordination. In contrast, Setoh, Scott, and Baillargeon (2016); (also Scott, 2017) argue that despite requiring elicited responses BCT’s task is easier than the standard false belief task because it lacks the need for inhibiting a prevalent (reality oriented) response.

Evidently, BCT’s results are of great theoretical importance for the field. For this reason we decided to have a closer look at their replicability and interpretation. Although there are quite a few demonstrations of early sensitivity to belief, hardly any of them have been replicated by different laboratories (e.g., for BCT’s task Fiske, Butterfill, & Rakoczy, 2013 found similar results but they used a noticeably different procedure). Another question, of course, concerns the interpretation of the results. Again, although there are many demonstrations of infants’ sensitivity to belief in different situations, no study that we are aware of, has yet specifically tested the more recently suggested alternative interpretations for several of the VOE and AL studies (e.g., Ruffman, 2014; Wellman, 2014, chp 8). To our knowledge the only study that tested an alternative explanation for BCT’s findings was run by Allen (2015).

Allen (2015) contrasted the FB condition with a new *clairvoyance* condition, which was the same as the FB condition except that E2 tried to open box B, where the toy is but E2 thinks it is empty. If, so Allen reasoned, children take E2’s false belief into account E2 must be intending to open an empty box, so they should direct E2 to box A, which is empty. Children did not do this. They helped E2 find the toy as often in the clairvoyance as in the FB condition. However, Allen’s argument is not very persuasive, since trying to open a box of which one thinks that it is empty, does not mean one wants to open box A because one thinks it is empty. One might have plenty other reasons for opening it.

To gain experience with BCT’s paradigm we started with a straight replication of the original TB and FB conditions.² We soon noticed that with the FB-TB manipulation not only E2’s belief changed but the conditions had a quite different feel. One very obvious difference is the trick which E1 plays on E2 only in the FB condition. As already pointed out by Allen (2015, p.66): “*Hiding the toy in the context of playing a trick not only makes the toy particularly salient but it also creates an expectation that the adult is going to return and look for the toy*”. Two further potentially confounding factors pertain to ownership of the toy and to E2’s projected interest in the toy. The procedure of the FB condition (1) confirms the initial impression that the toy belongs

² We most gratefully acknowledge David Buttelmann’s personal visit to our lab in Salzburg and the intense exchange with both labs, Salzburg and Stirling, to instruct us in the details of how to conduct their study and to check the videos of our procedure.

to E2 because, for example, E1 only dares move it secretly in E2's absence. Children this age have already an acute sense of ownership (Nancekivell, Van de Vondervoort, & Friedman, 2013). (2) The condition also leaves unquestioned that E2 takes great interest in her toy. In contrast, E2's behavior in the TB condition signals that (a) it may not be E2's toy since he watches E1 move the toy without complaining that they have not asked for permission and (b) that E2's enthusiasm and interest for the toy must have waned towards the end since E2 watches as a bystander while E1 moves the toy. Ownership and interest therefore are further possible common sense reasons which could explain children's different behavior in the two conditions. Taken together we proposed three possible factors (i.e., ownership, interest and playing a trick, but there could be many more) that provide teleological reasons for children to show a distinct helping pattern in the two conditions that are not based on belief reasoning. However, the purpose of the study is neither to investigate the specific reasons that might motivate children's behavior nor the general nature of early pro-sociality. What we want to test is whether their responses in this task are based on belief reasoning or on some kind of teleological reasoning (as understood by Perner & Roessler, 2010).

Perner and colleagues (Perner & Esken, 2015; Perner & Roessler, 2010) argue that by 9–18 months children become 'teleologists' able to derive an agent's reason for an action without concern for the subjective views provided by mental states. The important feature of teleology is to see objective facts as providing the reasons for action. For instance: it starts to rain at the birthday party. The teleologist naturally perceives the need for a shelter for the birthday cake (the cake being under a shelter is preferable/more-desirable/better than it being left in the rain, which makes it a potential action goal) and uses her knowledge of where to find a shelter to bring the cake there. Although the evaluation of the cake being under a shelter as 'better' (or desirable) and the 'facts' of the shelter's location are based on the teleologist's subjective view, the teleologist treats these 'subjective facts' as objective. For the teleologist it is a simple fact that it is better to shelter the cake than leave it in the rain; it is not a matter of being considered better by some people depending on their subjective views of what is good or bad. This subjectivity is inherent in the mentalist concepts of *desire* (subjective view of what would be a better state, i.e., a goal) and *belief* (subjective view of what the facts are).

To illustrate how teleology applies to BCT's scenario we use the suggestion from above that children perceive the agent's (E2's) continued interest in playing with the toy. In the FB condition E2 is continuously playing with the toy (which makes her happy and thus for a pleasant, desirable situation). That E2 has to interrupt her play in order to fetch a key suggests that she will continue to play with the toy on her return. Hence, enabling her to do so will make for a better situation (she'll be happy; a goal to be achieved) than preventing her from doing so (she'll be nervous and grumpy). Consequently, when she is looking for the toy in the wrong box children have good reason to help her find the toy in box B to achieve a better situation (a goal). In the TB condition the agent interrupts her play with the toy and watches the child and E1 play with it and then briefly moves away to close a door. This gives less clear indication that the agent is likely to resume her play with the toy. So when she tries to open the empty box the teleologist child perceives no compelling reason to direct her to the box with the toy. To note: this explanation of the results by BCT does not depend on children making use of E2's belief about the content of the boxes to infer what E2 wants, as BCT suggested. There is no need for mentalizing; a basic concern about what purpose (goal) E2 pursues is directly indicated by her different behavior in the early stages of the TB and FB conditions.

With these assumptions outlined above BCT's data can be explained without children having to concern themselves with E2's beliefs or knowledge. When in the FB condition E2 returns it is very likely that she is coming back for her toy. When she is trying to open box A children recognize her error and correct her by redirecting her to her toy in box B. In the TB condition it is not so clear what E2 wants. When she tries to open box A she is likely to want to open this box – for whatever reasons – and thus children help her do so.

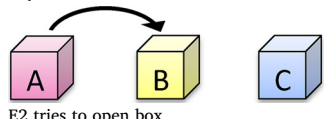
We tested our teleological alternative against BCT's mentalistic explanation of their data using three boxes, A, B, and C in four conditions. Two conditions, the *old-FB* and *old-TB* condition, corresponded to BCT's FB and TB conditions, except for the presence of the third box that was not used in the test procedure, only in the training phase. The *new-FB* condition conformed the original FB condition except that E2 tried to open the third and empty box C. In the *new-TB* condition E2 behaved as in the original TB condition except for trying to open the empty box C rather than the empty box A, into which E2 had originally placed the toy.

We expect that the presence of the third box will not influence children's helping behavior and that we will replicate BCT's original results in the *old-FB* and *old-TB* conditions. For the *new-FB* condition the two theories make different predictions. If children use E2's belief and knowledge to infer what E2 wants (BCT's theory) then children should behave in the *new-FB* condition in the same way as in the original TB condition: they should help open box C, since E2 knows that box C never contained her toy and she cannot be looking for it. Our teleological alternative predicts that since E2's behavior in the early phases of the *new-FB* condition was the same as in the original FB condition, this should signal directly that E2 is interested in getting her toy. So children should help her find it by directing her to box B that contains the toy. For the *new-TB* condition the two theories make the same behavioral prediction. BCT predict that children help open C because E2 knows that it is empty. The teleological alternative predicts that children help open C because E2, who lost interest in the toy, is now interested in opening box C. One should point out, though, that teleology by itself does not yield a strict prediction of how interested children should be in opening box C. Only because we know from the results of BCT that they are under these conditions that we make this prediction. Table 1 provides an overview of the four conditions and each theory's predictions.

2. Study 1

The objective of this study was not primarily to see whether BCT's result can be replicated but whether we would manage to

Table 1
Predictions of BCT's mentalizing hypothesis and our teleological alternative per condition.

Condition	Procedure	Predictions	
	<p>Toy was in A, is now in B, never was in C.</p> <div><p>E2 tries to open box</p></div>	BCT's mentalizing hypothesis	Our teleological alternative
		Children help with box	
Old-FB	A	B	B
Old-TB	A	A	A
New-FB	C	C	B
New-TB	C	C	C

replicate it using – within our possibilities – the same materials and procedure before we proceeded to test an alternative interpretation of the data in Experiment 2.

2.1. Method

2.1.1. Participants

Overall, 45 children between 18 and 32 months ($M_{age} = 24.47$, $SD = 4.08$, 20 girls) participated in the study. The age range was chosen to cover the range between BCT's youngest in their sample of 18 month olds in Study 2 and their oldest in Study 1. These two groups showed about the same size of effect, which we try to replicate. A third group of 16 month olds did not show a significant effect and so we did not include this age in our study. Data were collected in the Theory of mind Child Lab of the University of Salzburg ($n = 20$), the Parent-Toddler Group of the University of Stirling, ($n = 17$) and in the Little Stars Nursery ($n = 8$). Seventeen children had to be excluded because of parental/teacher error (3), fussiness (10), unclear responses (2), or because they did not respond to any helping request by opening or at least touching one of the boxes (2). Compared to BCT, who excluded 23% of the 2,5-year-olds and 54% of 18–16 month-olds, our overall dropout rate was similar (35%). For child centered reasons only (fussiness and no response) BCT excluded 6.5% of the 2,5 year olds and 26% of the 16–18 months olds while we excluded 13% of the older children (26–32 months) and 41% of the younger children (18–25 months olds) of our sample. The 28 children of the final sample had a mean age of $M_{age} = 25.64$ months, ($SD = 3.64$, range = 20–32 months, 12 girls).

2.1.2. Materials

Materials were produced according to the description in [Buttelmann et al. \(2009\)](#). Boxes were identical in size, virtually identical in locking mechanism, handlebars and color; for the stuffed toy we also used a caterpillar which was roughly the same size (48 cm). Following David Buttelmann's advice, the locking mechanisms of the boxes were loosened so that they would make a noise when E2 tries to open them and the caterpillar was stuffed especially with plastic foil so that it made sizzling noises when being moved. Both features may be important for grabbing children's attention during the procedure. For the warm up and instruction phase a wooden pegboard game was used.

2.1.3. Procedure

For a detailed description of the procedure see [Buttelmann et al. \(2009\)](#). To replicate the original procedure we prepared transcripts (English and German) of a video provided by David Buttelmann. We then videotaped our procedure and received written feedback from him for both, the English and the German version. We implemented his detailed feedback on the protocol as well as on our realization of the procedure.

As in BCT children participated either in a TB or in a FB condition. Position and color of the boxes as well as where the toy was placed first were counterbalanced. E2's eye gaze before attempting to open a box was also counterbalanced. Test-sessions were videotaped and coded by two independent raters. As in the original study, it was coded which box a child opened or touched first. Our coding and exclusion criteria followed closely that of BCT. However, in the FB condition two children showed an interesting response pattern that was not reported by BCT but also occurred in [Allen's \(2015\)](#) study with 3–5-year olds; they responded by opening the old toy box first but not to help E2 to open it but only to gleefully show E2 that the box is now empty. Immediately after doing so they opened the current toy box to show E2 that the caterpillar had been hidden there. Both children additionally verbalized "Here it is!" Although these children touched the old box first, we interpret this behavior in BCT's favor as an understanding that E2 did not know the new location of the caterpillar. In order to avoid false negatives a "re-coding" according to this interpretation was added to the "original coding" by BCT and we will report results for both. Overall, two raters disagreed only on two trials. As a result, one child was excluded due to an unclear response. For the other child agreement could be reached through discussion with a third rater.

2.2. Results & Discussion

The 28 children needed different amounts of prompting to either touch or open a box. Eleven children spontaneously responded

Table 2

Number of children's responses directed at one of two boxes in our Study 1 compared to responses in BCT's (2009) Experiment 1 (30 months) and Experiment 2 (18 months) combined.

Condition	Study 1			BCT (2009)		
	Response directed at box		Total n	Response directed at box		Total n
	A now empty	B with toy		A now empty	B with toy	
FB	1 ^a	13 ^a	14	9	28	37
TB	6	8	14	30	7	37

^a Two children who approached box A to show that it is empty and then fetched the toy from B have been recoded as directing their response to B.

to E2's nonverbal request, nine children responded to one of E1's prompts, four children responded to one of E2's verbal prompts and four children helped only when their parent or teacher prompted (3) or assisted (1) them.

The center panel of Table 2 shows the number of children choosing either box B with the toy (two of them re-coded in the FB condition) or the empty box A. The right panel shows the original results by BCT in comparison. As one can see the proportion of children helping to open box B that contained the toy in the FB condition in our study is very high (recoded: 93%, original coding: 79%) and comparable to the original study (76%) and differs from a uniform distribution (Binomial test: re-coded: $p = .002$; original: $p = .057$). In contrast, the distribution of responses in the TB condition looked different. The proportion of the expected empty-box response was not even half (43%) in our study as opposed to the original study (81%) and the responses did not differ significantly from a uniform distribution (Binomial test: $p = .79$ in our case). If we compare our results in the TB condition directly with BCT's the difference is highly significant ($\chi^2 = 7.15$, $p < .01$). We have no good explanation to offer for this difference except for the observation that most children appeared to be at a loss of what is being asked of them in this condition. Despite this different performance in the TB condition, we nevertheless find a trend of a different proportion of B (box with toy) responses in the two conditions: Fisher's exact test $p = .038$ (relying on the recoded responses, with original coding: $p = .21$, both one-tailed) but two-tailed $p = .077$ (relying on the recoded responses, with original coding: $p = .42$). Hence, the data do not clearly speak against the null-hypothesis. We will take up this point again in the General Discussion. Since our study supported children's strong tendency to help the agent in the FB condition of the original study we thought to be in a good enough position to venture a test of our alternative explanation for that condition.

3. Study 2

The prime objective of this study was to assess how children in the FB condition infer from E2's attempt to open an empty box that E2 must be looking for her toy. Is it critical for their inference that the agent is trying to open that box, in which she mistakenly thinks her toy is located or is their inference based on some other information? We have suggested at least three other factors that may lead to that inference: in the FB condition – in contrast to the TB condition – children are primed by the initial part that (1) the toy belongs to E2, (2) E2 has great interest in playing with her toy, and is likely to engage in a hide and seek routine (Allen 2015). Any one of these factors make children expect that E2 will look for her toy on her return. When E2 returns and tries to open the now empty box children help her find the toy, without any concern for her false belief about where the toy now is. Children do not show the same helping behavior in the TB condition since the TB condition strongly suggests that the toy does not belong to E2, that she has lost interest in the toy, and because no hide and seek routine is indicated. Hence when she tries to open the empty box fewer children assume that she is still looking for the toy and that she tries to open the empty box for some unknown reason. Consequently, fewer children direct her to the toy and, instead, help open the empty box.

A new version of the FB task (*new-FB*) with three boxes can distinguish between the different explanations. Instead of trying to open the box that she believes contains her toy, E2 tries to open the third box which is also empty and has never contained the toy. If children rely on E2's false belief to figure out that she is likely looking for her toy – as BCT suggested – then they should come to the conclusion that she must have some other reason for trying to open this box since she knows that it does not contain her toy. If, however, children assume that the agent is keen to play with her toy again, as we surmised, then they should also help her to find the toy in this *new-FB* condition.

In order to tie in the results with the 2-boxes version, there will be an *old-FB* and an *old-TB* condition identical to the two boxes experiment in Study 1, except for the presence of a third box (which is involved in the training but not in the transfer in the test phase). For these two conditions we anticipate to replicate the results from Study 1. What differs in the *new-TB* condition is that instead of attempting to open box A the agent tries to open the third box C (as in the *new-FB* condition). Both theories predict the same behavior, that is, children will show the same response pattern in the *new-TB* as in the *old-TB* condition. Its purpose is to check whether the active involvement of the third box creates any deviation from what is expected.

3.1. Method

3.1.1. Participants

Overall, 126 children between 18 and 32 months were tested either in the Theory of mind Child Lab of the University of Salzburg

($n = 20$), in different childcare institutes in the city of Salzburg ($n = 87$) and in Scotland ($n = 19$). Testing in institutes took place in a separate room and in the presence of the child's teacher or parent. Thirty-six children (28%) had to be excluded due to parental/teacher (4) or experimenter error (4), fussiness (20), unclear responses (3) or because they did not respond to any helping request (5). Overall, 29.1% of children (20.6% of the older (28–32 months) and 37.5% of the younger (18–27 months) ones) were excluded. The final sample consists of 90 children between 18.04 and 32.82 months ($M = 27.15$ months, $SD = 3.65$, 40 girls).

Thirty-seven children participated in the replication conditions ($M_{age} = 27.17$ months, $SD = 3.69$). Six children spontaneously responded to E2's nonverbal request, 13 children responded to E1's prompts, five children responded to E2's verbal prompts, 11 children responded to their parents/teachers prompt and one child needed parental/teacher assistance. Fifty-three children participated in the new conditions ($M_{age} = 27.13$ months, $SD = 3.66$). Seven children spontaneously responded to E2's nonverbal request, 26 children responded to one of E1's prompts, eight children responded to one of E2's verbal prompts, 10 children responded to their parents/teachers prompting and one child needed parental/teacher assistance (unfortunately, for one child video recording is missing and therefore amount of prompts cannot be reported).

3.1.2. Materials & Procedure

In this study we added a third, differently colored but otherwise identical, box. All remaining materials were the same as in Study 1. The three boxes were set up in a semi-circle in equal distance (80 cm) to each other. The distance between the child and each box was 1 m. As in Study 1 the experiment started with E2 discovering the boxes. She opened and closed the lids of all three boxes several times before leaving the room so E1 could explain the locking mechanism to the child. By doing that E1 treated all boxes identically and according to the protocol of Study 1. The procedure in the test-phase was also identical to Study 1, with the exception that E1 locked the third box with the pin without attracting the child's attention to it and that E2 was sitting in some distance behind the three boxes in the decision-phase. Further details can be found in the supplementary materials.

3.1.3. Design & Coding

Children participated in one of the four conditions. We had two TB conditions and two FB conditions that differed according to which of the three boxes E2 tried to open (see Table 1). In neither condition the third box C was involved in the transfer of the toy from box A (old) to box B (current). Two of the conditions (old-FB and old-TB) were a replication of the original study, with the only exception that the third box was also present in the setting. In these conditions, E2 always tried to open the now empty box A. In both of the new conditions (new-FB and new-TB), E2 always tried to open the third, non-involved box C. Colours of boxes were assigned to position according to a Latin Square Design within each condition. So the location of the boxes was fully counterbalanced, as was where the toy was put first. This makes for 36 different combinations to which children were randomly assigned without using a combination twice. The direction of transfer was varied such that the toy was always transferred from box A to the box to the right, if A was the box on the right the toy was transferred to the box on the left. Eye gaze was varied such that E2 always looked at box B first, then at the left empty box (which could either be box A or C), then at the right empty box (which could again either be box C or A). Test-sessions were videotaped and coded by two raters. Again, it was coded which box a child opened or touched first. Two raters agreed on 87 of 90 trials and the disagreements were resolved through discussion with a third rater. Again, three children in the FB conditions showed a response as described in Study 1: after opening the box that E2 tried to open they showed her that it was empty and immediately proceeded to show her where the caterpillar was hidden. We re-coded these children as directing their response (primarily) at box B. We will again report results according to this re-coding as well as for the original coding.

3.2. Results & Discussion

Table 3 shows children's responses in the four conditions. Children practically ignored the third box C (with the exception of one child) and showed a preference for box B, containing the toy, in both conditions but not significant in either (Binomial Test both $ps > .11$). However, the pattern of results for A-responses and B-responses is not significantly different from the results of Study 1 ($\chi^2(3) = 3.212, p > .36$).

The results of the old FB and TB conditions, which we hoped would replicate the results of the two box conditions in Study 1, indicate that the use of three boxes flattened the distribution of chosen boxes. This suggests increased error responding. Closer inspection of the data showed that children had a strong preference for approaching the center box, the one closest to E2. The choices

Table 3

Number of children's responses directed at one of the three boxes (re-coded).

Condition	Agent tries to open	Response directed at box			Total n
		A: box now empty	B: box with toy	C: box always empty	
Old-FB	Box now	6 ^a	12 ^a	1	19
Old-TB	empty	7	11	0	18
New-FB	Box always	2	18 ^a	6 ^a	26
New-TB	empty	7	8	12	27

N.B.: Numbers circled are children's responses directed at the box the agent tries to open. ^aOne child was re-coded from box A to box B in the old-FB and two children from box C to B in the new-FB condition.

Table 4

Number of children with center box responses excluded (re-coded).

Condition	Agent tries to open	Response directed at box			Total n
		A: box now empty	B: box with toy	C: box always empty	
Old-FB	Box now	2 ^a	6	0	8
Old-TB	empty	6	8	0	14
New-FB	Box always	1	9	2 ^a	12
New-TB	empty	2	2	4	8

N.B.: Numbers circled are children's responses directed at the box the agent tries to open. ^aOne child was re-coded from box A to box B in the old-FB and one child from box C to B in the new-FB condition.

of boxes for all four conditions were 26% left, 53% center, and 21% right box. This differed from the expected uniform distribution ($\chi^2(2) = 9.8, p > .007$), which should have occurred since assignment of A, B, and C to box location was counterbalanced. Evidently the center box attracted children, which increased error trials. So we excluded this source of error by looking only at responses directed at the left or right box, shown in Table 4. Now the picture becomes much more accentuated. The response frequencies for the old conditions now resemble closely those of Study 1. Without the distraction of the center box Study 2 replicates Study 1, which is reassuring as it shows that the use of the third box does increase error but does not distort the results.

We now turn to our core concern, children's responses in the new-FB condition with our focus on children's response directed at the box with the toy (B) in relation to the box, which E2 tries to open (C). According to BCT's theory children will know that E2 believes that the toy is in box A and should therefore show the same preference for the box E2 is trying to open (C), as they do in the original TB condition. Our theory predicts, however, that children assume that E2 is looking for her toy and, therefore, direct her to the box with the toy (B), as they do in the original FB condition. The data confirm that more responses were directed at B than at C. To see whether this refutes one and supports the other theory we calculated a Bayes Factor (BF). Each theory is tested against the null hypothesis H_0 of no preference between the two boxes. We use Rouder's Bayes calculator for binomial data: <http://pcl.missouri.edu/bf-binomial>. For specifying the model of H_1 for BCT's theory we use the observed proportions from BCT's original TB condition (see Table 2) since their theory lets us expect similar results to their TB condition in our new-FB condition. The model thus specifies the proportion of 30/37 for C (which the agent tries to open) vs 7/37 for B (where the toy is). The corresponding observed proportions in the new-FB condition of 6/24 for C and 18/24 for B result in $BF = 0.00076$, very strong evidence against BCT's theory³ (Dienes, 2014; Lee & Wagenmakers, 2014). In contrast, our explanation predicts that children should behave in the new-FB condition as in BCT's FB condition. Thus we use the data from their FB condition with 28/37 for B (where the toy is) vs. 9/37 for C (which the agent tries to open) to specify H_1 . Given the observed proportions this yields $BF = 17.86$ substantial evidence for our hypothesis. The same result is obtained for the reduced data set excluding middle box responses (see Table 4) with $BF = .005$ for BCT's theory and $BF = 9.09$ for our theory.

Moreover, there were significantly more B responses in the new-FB condition than in the new-TB condition in relation to C responses (Fisher's Exact $p = .031$). One can argue that what matters for BCT's mentalist claim is that the original FB condition should lead to more B-directed responses because E2 is trying to open the box she mistakenly thinks her toy is in. Whether children direct their helping at the box E2 is trying to open in the TB condition is not really of essence. In fact our data suggest that children do not have much idea of what they are supposed to do in the TB condition. Hence a fairer test might be to contrast the number of B-directed responses with responses directed at any one of the other boxes (A or C). In this case we get a highly significant difference between conditions: Fisher's exact test $p = .006$.⁴

4 General discussion

The overarching aim of this paper was to test an alternative interpretation of Buttelmann et al. (2009). Study 1 was a direct replication attempt to build a fundament for testing this alternative hypothesis. Like in the original study, children's responses in the FB condition were much more often directed at the box that contained the toy than at the empty box. We could also show that this tendency was stronger in the FB than in the TB condition, albeit children not showing the expected preference for the empty box in the TB condition in our study.

The theoretical thrust of our data comes from our new-FB condition in Study 2 designed to distinguish BCT's original mentalistic explanation from our teleological alternative. BCT claimed that in order to be able to help the agent children must infer in the FB condition that the agent wants her toy as she is trying to open the box where she *believes* her toy to be. Whereas in the TB condition, the agent tries to open a box, which she knows is empty, she probably wants to open that box for unknown reasons. In our new-FB condition the agent tries to open box C, which she knows to be empty. Therefore, on BCT's reasoning, she cannot be looking for her toy but must intend to open that box for some other reason. Consequently children should direct their response to this box the agent is trying to open, in analogy to the original TB condition. Our data speak strongly against this explanation.

³ $BF < 1/3 = .33$ is considered substantial but moderate, $BF < 1/10 = .10$ strong evidence for H_0 and against H_1 . Conversely, $BF > 3$ counts as substantial but moderate, and $BF > 10$ as strong evidence for H_1 .

⁴ The same test for our reduced data set (excluding middle box responses): $p = .04$ one-tailed.

Our suggestion is that children infer what the agent wants without concern for her mental states. The procedure in the FB condition signals that she is still highly interested in her toy or is likely to engage in a hide and seek routine when she returns, therefore children will help her find her toy when she makes an error of looking into box A instead of box B. For our new-FB condition this theory, as opposed to BCT's theory, predicts the same behavior as for the old-FB condition since the pre-test procedure is exactly the same. Children – according to our hypothesis – are fairly sure that the agent is coming back to look for her toy. However, she goes to the wrong box (A in the original and C in the new-FB condition), which provides children with a good reason to help her find the toy in box B. The data speak for this explanation much more strongly than for the one advanced by BCT.

Our explanation has potential relevance for how we should look at infants' impressive social competence. An important issue concerns the question of why such young children are so keen to help others. Paulus (2014) outlined several different classes of models of what motivates pro-social behavior in very young children. (1) Emotion-sharing-models assume that helping behavior arises as a result of emotional contagion in combination with the development of self-other-differentiation and the arising ability to respond to others' negative emotions in a solution oriented manner, for example comforting (Hoffman, 2000; Preston & De Waal, 2002). (2) Social-interaction-models propose that children act pro-socially merely because they enjoy interacting with other people without a specific motivation to be of benefit to others (Over & Carpenter, 2009; Reingold & Merikle, 1993). (3) The social-normative-model emphasizes the role of the social environment and characterizes the emergence of helping behavior as a process of internalizing the rules of their environment (in support of this model see e.g., Hammond & Carpendale, 2012). (4) Proponents of goal-alignment-models (e.g., Kärtner, Keller, & Chaudhary, 2010; Kenward & Gredebäck, 2013) agree upon the idea of a goal contagion process by which children take over the other's goal and consequently act as if it was their own. Goal-alignment-models are the most similar to the (5) teleological account as both propose that an understanding of goals, and not an understanding of others' mental states, is the driving factor for helping. In Perner and Roessler's account, however, there is no need for a process of goal contagion because teleological reasoning is based on objective facts and seeing the possibility of a desirable state should give anyone reason to make this the goal of their action.

Although these models differ about children's motivation for engaging in helping behavior, they all presuppose that children in our study know what the purpose/goal of the other person's action is. On all five accounts children need to know this in order to (1) respond adequately to other's emotional state, to (2) engage in behavior apt to promote good interaction, to (3) show that one is willing to help, or to (4) take on the other's goal. The central question tested in our new conditions is whether children need a belief-desire theory to do so, or whether they can do it on the basis of what they observe the agent doing. This question is most explicitly addressed in teleology which is particularly explicit that no mental states are needed. This stands in stark contrast to how Tomasello (2014) describes the cognitive basis of cooperation and helping.

Tomasello (2014) argues that an early inclination for helping stems from a specific human genetic trait, a cooperative turn in human evolution, which consists of the ability to engage in higher order mentalizing. For instance in the object choice task, where children or apes are faced with several up-side down buckets, one of which is baited, when the experimenter marks or points to one of them, very young children spontaneously look for the bait under this bucket while chimpanzees need excessive training to learn this. According to Tomasello (2014, p. 57) *“the key point is that the inferences used in cooperative communication are socially recursive... In the object choice task..., the recipient infers that the communicator intends that she knows that the food is in that bucket – a socially recursive inference that great apes apparently do not make.”*

Our finding put into question whether such intricate, recursive mentalizing abilities are needed to explain young children's cooperative inclinations. But, if not recursive mentalizing, what does give children this cooperative knack? Roessler and Perner (2015) and Perner and Esken (2015) have argued that an understanding of the reasons for acting (teleology) provides a more direct, hence less vulnerable, basis for cooperation than recursive mentalizing. Reasons for action consist of non-mental, objective facts including value facts, that is, what is good or bad to have. Our explanation of BCT's results implements this approach even though we cannot be sure exactly how the children interpret the interactions. We have mentioned several possibilities. For instance, we proposed that children notice E2's interest and emotional engagement with the toy and therefore see her playing with it as desirable (a goal). Since it is a 'good' thing it provides reason – not just for the agent – for everyone, who can contribute to help bring it about. So when the agent does something inadequate for achieving this goal the child teleologist has a natural inclination to help—without any need to engage in reasoning about each other's desires and beliefs. Allen (2015) has pointed out another plausible possibility. The secretive cue in the FB condition may give children the impression that this is a game of hide and seek, in which the seeker is to find the hidden target. We know that children at this age tend to help by directing the seeker to the target. From an adult's point of view this helping is counterproductive and misses the point of the game; yet that is what children tend to do (Gratch, 1964; Theo Wimmer anecdote in Perner, 1991, p. 153). From a teleologist's point of view this counterproductive help makes perfect sense. The overall goal, when everybody is happy, thus a desirable, good state to be in, is for the seeker to find the target. So the teleologist child chips in to get to that goal by helping the seeker.

Our result has also implications for how we should view other supposed demonstrations of early belief understanding. Of these – there are many by now – only few, if any, have been replicated in a strict sense and some, as this journal issue attests, are not easily or perhaps not at all replicable. A recent number of studies report severe replication issues (see e.g., Dörrenberg, Liszkowski, & Rakoczy, 2017; Powell, Hobbs, Bardis, & Carey, 2017; Schuwerk, Prieuwater, Sodian, & Perner, 2017; Yott & Poulin-Dubois, 2016). Even the data of our Study 1 did not reach the same significance criterion (two-sided p-value below .05) as the original study. But one should not conclude from that that our data provide evidence against the existence of BCT's effect, that is the difference between the two conditions. To see this we carried out a Bayes analysis which shows a BF = 2.81 in favor of BCT's finding, but a BF below 3.0 is regarded as inconclusive evidence (Dienes, 2014; Lee & Wagenmakers, 2014).

Nevertheless, by the sheer weight of numbers of other kinds of demonstrations, early belief understanding is widely accepted as a

fact. For many of the findings alternative explanations have been proposed (e.g., Apperly & Butterfill, 2009; Fenici, 2015; Perner & Roessler, 2010; Ruffman 2014; Wellman, 2014, chp 8), but few of the pithier ones have been tested.⁵ Peter Carruthers (2013, p. 150), in a recent evaluation of the evidence, concluded with soothing caution: “...at present we seem warranted in tentatively endorsing the infant-mindreading hypothesis, based on its record so far,” and he added wisely: “But if it should turn out that these existing studies cannot be replicated, or if additional control experiments provide evidence of non-mentalizing mechanisms underlying the results, then the situation may yet reverse itself.” The present study heralds this reversal.

Acknowledgements

We are grateful to children and their parents who participated in this study. We thank the kindergarten teachers in the Krabbelgruppen Schneckenhaus, Krabbelgruppe Grisu, Krabbelgruppe Kindernest, Krabbelgruppe Vogelweiderstrasse, Kuschelkiste Morzg, Kinderparadies Mayrwies, Krabbelgruppe Sternschnuppe, Kuschelkiste Gneis, Krabbelgruppe Spieloase, Krabbelgruppe St. Johannis-Spital, English Playcorner, Krabbelgruppe Uni Freisaal in Salzburg, Hoppolino Anif as well as the Parent-Toddler-Group Stirling, Little Stars Stirling, Dunblane Nature Kindergarten and the Corstorphine Playgroup. We thank Eva Leinsle, Melanie Tockner, Sabine Schellin, Anna-Lena Borowski, Katharina Schweller, Bianca Schuster and Emmily Fedra as well as Joshua March, Megan Everts, Stephanie Gibb, Kalliopi Kremastioti, Åse Rading, Jenny Lundmark and Eszter Karath for collecting data. We further thank Stefanie Moser for help with script preparation and Nicole Wilson and Caragh O'Brien for help with data preprocessing. Special thanks are owed to Frank Esken and Matthias Schurz for inspiring the idea for the three boxes experiment in the initial discussion at the Summer School in Pärnu.

This research was partially funded by the Austrian Science Fund project (I637–G15) “Rule-Understanding, subjective preferences, and social display rules” as part of the ESF EUROCORES Programme EuroUnderstanding initiative.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.cogdev.2017.08.002>.

References

- Allen, J. W. P. (2015). How to help: can more active behavioral measures help transcend the infant false-belief debate? *New Ideas in Psychology*, 39, 63–72. <http://dx.doi.org/10.1016/j.newideapsych.2015.07.008>.
- Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review*, 116, 953–970.
- Baillargeon, R., Scott, R. M., & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Sciences*, 14(3), 110–118. <http://dx.doi.org/10.1016/j.tics.2009.12.006>.
- Buttelmann, D., Carpenter, M., & Tomasello, M. (2009). Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition*, 112(2), 337–342. <http://dx.doi.org/10.1016/j.cognition.2009.05.006>.
- Carruthers, P. (2013). Mindreading in infancy. *Mind & Language*, 28(2), 141–172.
- Clements, W. A., & Perner, J. (1994). Implicit understanding of belief. *Cognitive Development*, 9, 377–397. [http://dx.doi.org/10.1016/0885-2014\(94\)90012-4](http://dx.doi.org/10.1016/0885-2014(94)90012-4).
- Clements, W. A., & Perner, J. (2001). When actions really do speak louder than words-but only implicitly: young children's understanding of false belief in action. *British Journal of Developmental Psychology*, 19, 413–432. <http://dx.doi.org/10.1348/026151001166182>.
- Dörrenberg, S., Liszkowski, U., & Rakoczy, H. (2017). Does the infant really have a ToM? *Poster session presented at the international convention of psychological science [March]*.
- Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, 5, 781. <http://dx.doi.org/10.3389/fpsyg.2014.00781>.
- Fenici, M. (2015). Social cognitive abilities in infancy: Is mindreading the best explanation? *Philosophical Psychology*, 28, 387–411.
- Fizke, E., Butterfill, S. A., & Rakoczy, H. (2013). Toddlers' understanding of false belief about an object's identity. *Poster presented at the meeting of the society for research in child development. Seattle, USA*.
- Gratch, G. (1964). Response alternation in children: A developmental study of orientations to uncertainty. *Vita Humana*, 7, 49–60.
- Hammond, S. I., & Carpendale, J. I. M. (2012). Rethinking children's everyday help: Often active, occasionally unhelpful, and guided by parents. *Poster Presented at the Biennial Meeting of the International Society on Infant Studies*.
- Helming, K. A., Strickland, B., & Jacob, P. (2014). Making sense of early false-belief understanding. *Trends in Cognitive Sciences*, 18(4), 167–170. <http://dx.doi.org/10.1016/j.tics.2014.01.005>.
- Helming, K. A., Strickland, B., & Jacob, P. (2016). Solving the puzzle about early belief-ascription: Solving the puzzle about early belief-ascription. *Mind & Language*, 31(4), 438–469. <http://dx.doi.org/10.1111/mila.12114>.
- Hoffman, M. L. (2000). *Empathy and moral development. Implications for caring and justice*. Cambridge, UK: Cambridge University Press <http://dx.doi.org/10.1017/CBO9780511805851>.
- Kampis, D., Parise, E., Csibra, G., & Kovács, Á. M. (2015). Neural signatures for sustaining object representations attributed to others in preverbal human infants. *Proceedings of the Royal Society B: Biological Sciences*, 282(1819), <http://dx.doi.org/10.1098/rspb.2015.1683>.
- Kärtner, J., Keller, H., & Chaudhary, N. (2010). Cognitive and social influences on early prosocial behavior in two sociocultural contexts. *Developmental Psychology*, 46, 905–914. <http://dx.doi.org/10.1037/a0019718>.
- Kenward, B., & Gredebäck, G. (2013). Infants help a non-human agent. *PLoS ONE*, 8(9), <http://dx.doi.org/10.1371/journal.pone.0075130> e75130.
- Kovacs, A. M., Teglas, E., & Endress, A. D. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science (New York, N.Y.)*, 330(6012), 1830–1834. <http://dx.doi.org/10.1126/science.1190792>.

⁵ One exception is the study by Träuble, Marinovic, and Pauen (2010) who tested the suggestion by Perner and Ruffman (2005) that children use behavior rules like “agents tend to look for objects where they last saw them.” Their negative evidence, however, interpreted behavior rules very narrowly for seeing only and not for touching or other perceptual contact which one can also perceive and need not infer (see Povinelli & Vonk, 2004; Perner, 2010). Another is a study by Sodian and Thoermer (2008) replicating the finding by Onishi and Baillargeon (2005) but also showing that children's responses are more likely due to understanding absence of information rather than false belief. Allen's (2015) study on BCT's helping paradigm did not aim to replicate but tested an alternative interpretation on much older children than BCT.

- Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge University Press: Cambridge.
- Nancekivell, S. E., Van de Vondervoort, J. W., & Friedman, O. (2013). Young children's understanding of ownership. *Child Development Perspectives*, 7(4), 243–247. <http://dx.doi.org/10.1111/cdep.12049>.
- Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, 308, 255–258. <http://dx.doi.org/10.1126/science.1107621>.
- Paulus, M. (2014). The emergence of prosocial behavior: Why do infants and toddlers help, comfort, and share? *Child Development Perspectives*, 8, 77–81. <http://dx.doi.org/10.1111/cdep.12066>.
- Perner, J., & Esken, F. (2015). Evolution of human cooperation in Homo heidelbergensis: Teleology versus mentalism. *Developmental Review*, 38, 69–88. <http://dx.doi.org/10.1016/j.dr.2015.07.005>.
- Perner, J., & Roessler, J. (2010). Teleology and causal reasoning in children's theory of mind. In J. Aguilar, & A. Buckareff (Eds.), *Causing human action: New perspectives on the causal theory of action* (pp. 199–228). Cambridge, MA: Bradford Book, The MIT Press.
- Perner, J. (1991). *Understanding the representational mind*. Boston, MA: MIT Press.
- Perner, J., & Ruffman, T. (2005). Infant's insight into the mind: How deep? *Science*, 308, 214–216. <http://dx.doi.org/10.1126/science.1111656>.
- Powell, L., Hobbs, K., Bards, A., & Carey, S. (2017). Investigating the reproducibility of implicit theory of mind tasks with different levels of representational demand. In H. Rakoczy, & L. Kulke (Eds.), *Are implicit theory of mind findings robust? Some doubts from converging non-replications across the lifespan. Symposium conducted at the SRCD Biennial Meeting*.
- Preston, S., & De Waal, F. (2002). Empathy: its ultimate and proximate bases. *Behavioral and Brain Sciences*, 25(1), 1–20. <http://dx.doi.org/10.1017/S0140525X02000018>.
- Reingold, E. M., & Merikle, P. M. (1993). Theory and measurement in the study of unconscious processes. In M. Davies, & G. W. Humphreys (Vol. Eds.), *Consciousness: Vol. 1-Book, section*, (pp. 40–57). Oxford: Blackwell.
- Roessler, J., & Perner, J. (2015). Pro-social cognition: helping, practical reasons, and 'theory of mind'. *Phenomenology and the Cognitive Sciences*, 14(4), 755–767. <http://dx.doi.org/10.1007/s11097-015-9438-7>.
- Ruffman, T. (2014). To belief or not belief: Children's theory of mind. *Developmental Review*, 34(3), 265–293. <http://dx.doi.org/10.1016/j.dr.2014.04.001>.
- Schuwert, T., Priewasser, B., Sodian, B., & Perner, J. (2017). The generalizability and robustness of spontaneous false belief tests. In H. Rakoczy, & L. Kulke (Eds.), *Are implicit theory of mind findings robust? Some doubts from converging non-replications across the lifespan. Symposium conducted at the SRCD Biennial Meeting*.
- Scott, R. M. (2017). The developmental origins of false-belief understanding. *Current Directions in Psychological Science*, 26(1), 68–74. <http://dx.doi.org/10.1177/0963721416673174>.
- Setoh, P., Scott, R. M., & Baillargeon, R. (2016). Two-and-a-half-year-olds succeed at a traditional false-belief task with reduced processing demands. *Proceedings of the National Academy of Sciences*, 113(47), 13360–13365. <http://dx.doi.org/10.1073/pnas.1609203113>.
- Sodian, B., & Thoermer, C. (2008). Precursors to a theory of mind in infancy: Perspectives for research on autism. *The Quarterly Journal of Experimental Psychology*, 61, 27–39. <http://dx.doi.org/10.1080/17470210701508681>.
- Southgate, V., & Vernetti, A. (2014). Belief-based action prediction in preverbal infants. *Cognition*, 130. <http://dx.doi.org/10.1016/j.cognition.2013.08.008>.
- Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science*, 18, 586–592. <http://dx.doi.org/10.1111/j.1467-9280.2007.01944.x>.
- Surian, L., Caldi, S., & Sperber, D. (2007). Attribution of beliefs by 13-month-old infants. *Psychological Science*, 18, 580–586. <http://dx.doi.org/10.1111/j.1467-9280.2007.01943.x>.
- Thoermer, C., Sodian, B., Vuori, M., Perst, H., & Kristen, S. (2012). Continuity from an implicit to an explicit understanding of false belief from infancy to preschool age: False belief from infancy to preschool. *British Journal of Developmental Psychology*, 30(1), 172–187. <http://dx.doi.org/10.1111/j.2044-835x.2011.02067.x>.
- Tomasello, M. (2014). *A natural history of human thinking*. Cambridge, MA: Harvard University Press.
- Träuble, B., Marinovic, V., & Pauen, S. (2010). Early theory of mind competencies: Do infants understand others' beliefs? *Infancy*, 15(4), 434–444. <http://dx.doi.org/10.1111/j.1532-7078.2009.00025.x>.
- Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory of mind development: The truth about false belief. *Child Development*, 72, 655–684. <http://dx.doi.org/10.1111/1467-8624.00304>.
- Wellman, H. M. (2014). *Making minds: How theory of mind develops*. University of Michigan.
- Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13, 103–128. [http://dx.doi.org/10.1016/0010-0277\(83\)90004-5](http://dx.doi.org/10.1016/0010-0277(83)90004-5).
- Yott, J., & Poulin-Dubois, D. (2016). Are infants' Theory-of-Mind abilities well integrated? Implicit understanding of intentions, desires, and beliefs. *Journal of Cognition and Development*, 17(5), 683–698. <http://dx.doi.org/10.1080/15248372.2015.1086771>.