Journal of Experimental Child Psychology 204 (2021) 105058

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp

# Extended difficulties with counterfactuals persist in reasoning with false beliefs: Evidence for teleology-in-perspective

Eva Rafetseder [a,b,*], Christine O'Brien [c], Brian Leahy [d], Josef Perner [c,e]

[a] Faculty of Natural Sciences, University of Stirling, Stirling FK9 4LA, UK
[b] Department of Philosophy, University of Konstanz, 78457 Konstanz, Germany
[c] Department of Psychology, University of Salzburg, 5020 Salzburg, Austria
[d] Department of Psychology, Harvard University, Cambridge, MA 02138, USA
[e] Centre for Neurocognitive Research, University of Salzburg, 5020 Salzburg, Austria

ARTICLE INFO

*Article history:*
Received 12 November 2019
Revised 6 November 2020

*Keywords:*
Teleology-in-perspective
Counterfactual reasoning
False belief
Adaptive modeling
Theory theory
Simulation theory

ABSTRACT

Increasing evidence suggests that counterfactual reasoning is involved in false belief reasoning. Because existing work is correlational, we developed a manipulation that revealed a signature of counterfactual reasoning in participants' answers to false belief questions. In two experiments, we tested 3- to 14-year-olds and found high positive correlations ($r$ = .56 and $r$ = .73) between counterfactual and false belief questions. Children were very likely to respond to both questions with the same answer, also committing the same type of error. We discuss different theories and their ability to account for each aspect of our findings and conclude that reasoning about others' beliefs and actions requires similar cognitive processes as using counterfactual suppositions. Our findings question the explanatory power of the traditional frameworks, theory theory and simulation theory, in favor of views that explicitly provide for a relationship between false belief reasoning and counterfactual reasoning.



* Corresponding author at: Faculty of Natural Sciences, University of Stirling, Stirling FK9 4LA, UK.
*E-mail address:* eva.rafetseder@stir.ac.uk (E. Rafetseder).

https://doi.org/10.1016/j.jecp.2020.105058
0022-0965/

## Introduction

Counterfactual situations reflect the world as it would be had things been different. False beliefs are counterfactual insofar as they represent the world as it is not. Suppose that "Max" puts his chocolate into the drawer. Later, in his absence, his mum (mother) bakes a cake, uses some of the chocolate, and puts it in the cupboard. At this point, Max falsely believes that his chocolate is still in the drawer. Children older than 4 years typically predict that Max will search for his chocolate in the drawer even though it is no longer there. Younger children, until about 3½ years, indicate the item's true location (Wellman, Cross, & Watson, 2001).[1]

The false belief task has become an important indicator of children's acquisition of our folk psychology explaining how people act and why, which is thought to be based on mental states, in particular beliefs and desires. The task is for methodological reasons the best indicator of understanding belief as a mental state because it obligates a separation between the objective conditions and the agent's subjective view. Children's performance on the false belief task has been found to correlate with their ability to answer counterfactual questions (Riggs, Peterson, Robinson, & Mitchell, 1998; see also many subsequent studies in Fig. 1) around 4 years of age. This relationship remains difficult to explain for the traditional theories about folk psychology, for example, theory theory and simulation theory. For newer second-person perspective theories (e.g., embodied interaction theory: Gallagher, 2015; Gallagher & Hutto, 2008; pluralistic approaches: Fiebich & Coltheart, 2015; Musholt, 2018), it would still pose a challenge to incorporate counterfactuality into their account of the false belief task. By contrast, it has been suggested that teleology needs to make use of counterfactual reasoning in order to cover false belief cases (Perner & Roessler, 2010). That children should give the same answer to the belief question as to the counterfactual question follows directly from teleology-in-perspective. Straight teleology demands that people do and believe what they have objective reasons to do and believe. Objectively, Max should go to where his chocolate actually is. This, of course, would yield the reality error. So, teleology meets its limits unless one turns to using it within Max's perspective (teleology-in-perspective). This is possible by reasoning with a counterfactual conditional using Max's experiential record (i.e., all events minus the ones Max did not experience) as antecedent. Because Max has not seen that *the chocolate was moved,* one can reason counterfactually, "*If the chocolate had not been moved,* then the chocolate would be in the drawer." Hence, Max would have good reason to believe that his chocolate is in the drawer and to go to the drawer. And that is what a teleologist, using Max's perspective, should give for an answer.

Thus, the belief–counterfactual relationship promises differentiation between the different accounts. To sharpen this developmental relationship as a test of competing theories, we investigated its robustness. Research on children's counterfactual reasoning has shown that in certain tasks children have severe problems finding the correct answer into their teens (Rafetseder, Cristi-Vargas, & Perner, 2010). If the belief–counterfactual connection is robust, then belief reasoning in these tasks should be as difficult as counterfactual reasoning far beyond 4 years as traditionally thought.

Existing evidence pertains to 3- and 4-year-olds. Their answers to false belief questions ("Where will Max look for his chocolate?") were found to correlate with their answers to counterfactual questions ("If mother had not baked a cake, where would the chocolate be?") regardless of age and language skills (Riggs et al., 1998). The correlation remains stable even when executive functioning and processing load are controlled (see Fig. 1; other studies reported high correlations [e.g., German & Nichols, 2003, and Rasga, Quelhas & Byrne, 2016] but were not included in this graph because they had not controlled for age, language skills, or executive functioning). Brain regions that are active dur-

[1] Even though 3½-year-olds fail on direct questions, indirect measures such as anticipatory looking (Clements & Perner, 1994; Southgate, Senju, & Csibra, 2007; Thoermer, Sodian, Vuori, Perst, & Kristen, 2012) and violation of expectation (Onishi & Baillargeon, 2005; Surian, Caldi, & Sperber, 2007) demonstrate sensitivity to others' beliefs in infancy. There is still a debate about the cognitive basis of this early evidence (Ruffman, 2014; Setoh, Scott, & Baillargeon, 2016; Wellman, 2014). Children's looking behavior may reflect a sophisticated ability to track belief-like states (Low, Apperly, Butterfill, & Rakoczy, 2016) or to apply behavior rules (Perner, 2005). Recently, Kulke and Rakoczy (2018) and a special issue of *Cognitive Development* (Paulus & Sabbagh, 2018) reported severe replication difficulties, which suggests a qualitative difference between this early evidence and the changes observed around 4 years of age.

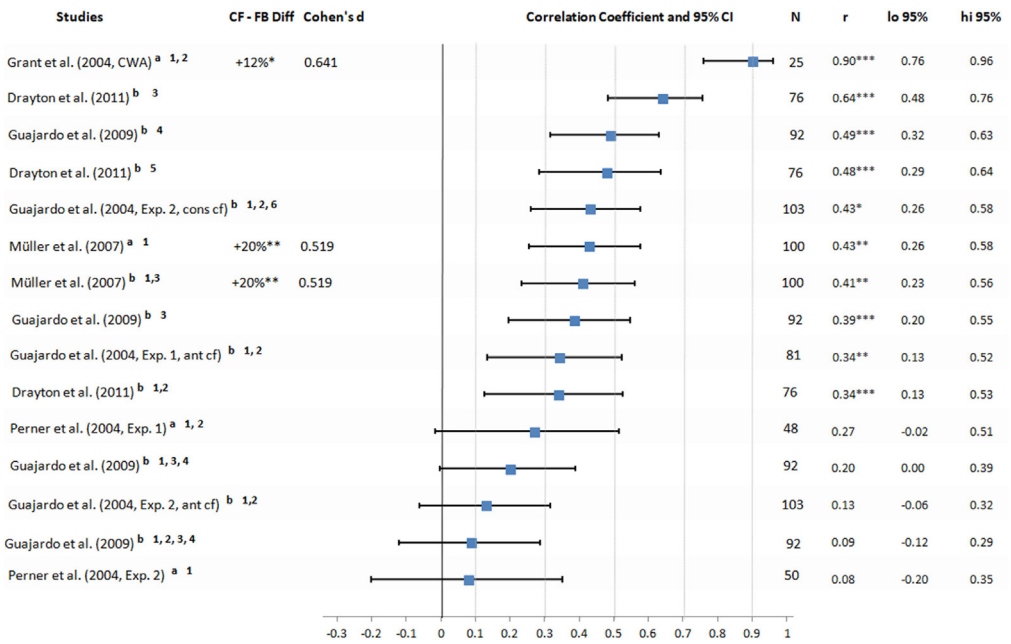


**Fig. 1.** Forest plot for 15 estimates of [a]partial correlations or [b]standardized beta coefficients from six studies controlling for [1]age, [2]language/verbal IQ, [3]working memory, [4]representational flexibility, [5]inhibitory control, and [6]antecedent counterfactuals. CF–FB Diff shows the mean difference between correct responses to counterfactual (CF) minus false belief (FB) questions. Cohen's *d* was corrected for dependence among means using Morris and DeShon's (2002) Equation 8. Studies are plotted in order of the size of the correlation. Horizontal lines represent the 95% confidence interval (CI), with lo 95% and hi 95% showing the lower and upper bounds. Asterisks represent the significance level of the correlation: $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$. Some data could not be included in the graph. Riggs et al. (1998), Experiments 2 and 3 (both controlling for age and verbal ability), did not report the beta coefficient, but counterfactual questions accounted for a portion of the variance [Experiment 2: $R^2(24) = .80$, $p < .001$; Experiment 3: $R^2(20) = .47$, $p = .005$]. Peterson and Bowler (2000) reported that counterfactual questions significantly predicted the false belief score beyond verbal mental age in typically developing children ($\chi^2 = 8.33$, $p < .05$) and in children with autism spectrum disorder ($\chi^2 = 7.31$, $p < .05$). Eddy, Beck, Mitchell, Praamstra, and Pall (2013) did not report correlation coefficients. (See above-mentioned references for further information.)

ing counterfactual and false belief tests also overlap to a large degree (Van Hoeck et al., 2014). Taken together, understanding of counterfactual situations seems highly predictive of performance on false belief tests.

Unfortunately, existing evidence for the relationship between counterfactual and false belief reasoning is purely correlational. Interpreting the results is hampered by lack of clarity about causal relationships. Here we report two experiments that manipulated the difficulty of the required counterfactual reasoning and checked whether the difficulty of the false belief test varied accordingly. Following established findings (Rafetseder et al., 2010; Rafetseder, Schwitalla, & Perner, 2013), we manipulate the difficulty of counterfactual reasoning tasks by manipulating the demands of the *nearest possible world constraint.*

### *The nearest possible world constraint*

When adults answer counterfactual questions such as "If mother had not baked a cake, where would the chocolate be?", they do not imagine counterfactual scenarios that depart gratuitously from the actual world (Edgington, 2004, 2008; Lewis, 1973; Stalnaker, 1968). They do not invent scenarios where the mother eats the chocolate anyway or where the chocolate floats over to the cupboard of its own accord or disappears magically. They imagine scenarios that are as similar as possible to the

actual world but where the counterfactual antecedent is true. All causal consequences of changing the antecedent are changed accordingly. But everything causally independent of the antecedent is unchanged.

On occasion, 3-year-olds answer as if they followed this constraint. For example, Harris, German, and Mills (1996) showed participants a puppet who wore her dirty shoes across a clean floor, leaving a trail of footprints. When asked, "If she had taken her shoes off, would the floor be clean or dirty?", even most 3-year-olds said "clean." By contrast, Rafetseder et al. (2013) showed participants two puppets, "Carol" and "Ben," both of whom left footprints on the clean floor. When asked, "If Carol had taken her dirty shoes off, would the floor be clean or dirty?", few children said "dirty" before 6 years of age.

The nearest possible world constraint demands that Ben still dirties the floor in the imagined scenario. This is because Ben wearing his shoes is causally independent of whether Carol takes off her shoes. Young children struggle to attend to these causal independencies when reasoning with counterfactuals (Nyhout, Henke, & Ganea, 2019). They use a simpler strategy for constructing counterfactual situations that we call *basic reasoning with counterfactuals* (BRC) here. Older children use a more sophisticated strategy that we call *mature reasoning with counterfactuals* (MRC) here.[2] These strategies are described in detail in Leahy, Rafetseder, and Perner (2014). BRC does not require that all causally independent features of the actual scenario are held fixed in the imagined scenario. So, BRC leads to hallmark errors when causally independent features of the actual world are relevant such as when Ben independently dirties the floor. We call these errors *BRC errors.*

Studies of the relationship between counterfactual reasoning and belief reasoning have not included conditions designed to reveal BRC errors. But if counterfactual reasoning is necessary for belief reasoning, then a properly designed false belief test should reveal children's systematic violations of the nearest possible world constraint. We developed two such tests. We adapted the procedure of Rafetseder et al. (2010) so that we could ask both a counterfactual question and a false belief question about the same sequence of events (see Fig. 2) and where it is possible to tell when children make the BRC error.

## Experiment 1

We modeled our test material on Rafetseder et al.'s (2010) study of counterfactual reasoning. It involved a dollhouse where a tall boy and a short girl live with their mother. Their mum often puts sweets for the children into the kitchen cupboard. She sometimes puts them on its high shelf, which only the tall boy but not his short sister could reach, or on the bottom of the cupboard, from where both children could get them. Whenever one of the children finds sweets, the child takes them to his or her own bedroom.

In the *discriminating condition,* a story went as follows. Today mum has put sweets on the high shelf. When asked a control question, "If the little girl comes, where will the sweets be?", even 3- to 5-year-olds answered correctly, "on the high shelf," by pointing out that she is too short to reach them, so they will stay on the high shelf. Then the tall boy came and took the sweets to his room. When asked, "If the little girl had come instead of the tall boy, where would the sweets be?", the very same 3- to 5-year-olds indicated the girl's room, the BRC error where they neglect the causally independent fact that the sweets had been on the high shelf where she cannot reach them. From around 6 years of age, children started to answer that the sweets would remain on the high shelf, the MRC answer that is also provided by most adults. Our focus was on children past 5 years of age, when nearly all master the traditional false belief test, to see whether their BRC errors on the counterfactual questions also occur on false belief questions. Therefore, we tested children aged 7 to 14 years and a

[2] We are making a break here with the terminology of our previous work. "MRC" replaces our term "CFR," which was short for counterfactual reasoning. The term was problematic because "counterfactual reasoning" was both our special term of art and a common term in the literature that referred to any strategy for reasoning with counterfactuals. Our terminological change eliminates this ambiguity. We adjust our earlier term "BCR"—for "basic conditional reasoning"—to capture the parallels between BRC and MRC.

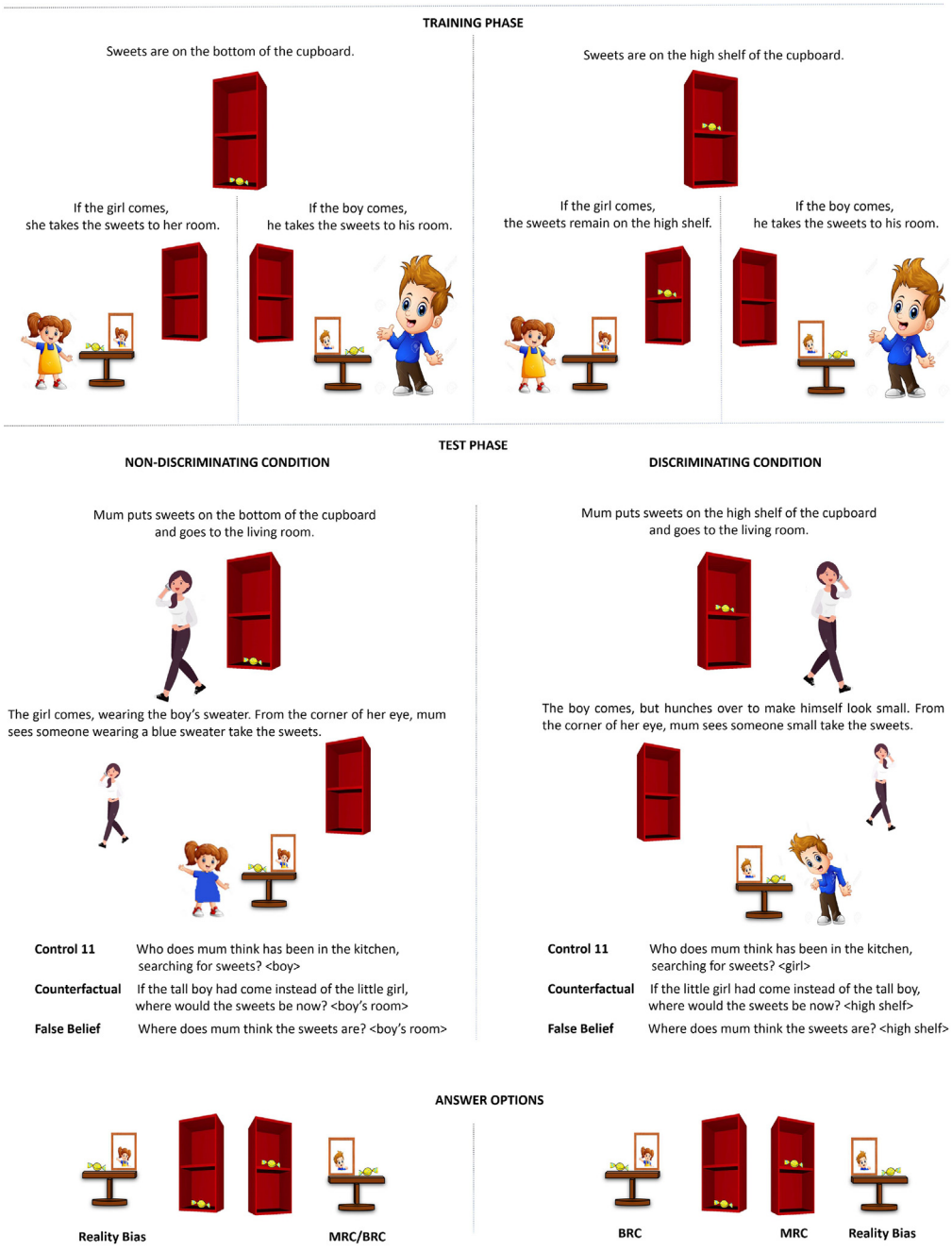


**Fig. 2.** Experimental setup in Experiment 1 split for nondiscriminating (left) and discriminating (right) conditions. Bottom shows answer options and processes leading to these answers. MRC, mature reasoning with counterfactuals; BRC, basic reasoning with counterfactuals.

group of adults (to see whether they conformed to our judgment of what constitutes a correct MRC answer).

In the *nondiscriminating condition,* mum put the sweets on the bottom of the cupboard and the girl took them to her room. The corresponding counterfactual question, "Where would the sweets be if the tall boy had come instead of the little girl?" was now easy even for the youngest children because the correct answer "boy's room" resulted from using BRC as much as from using MRC. Whether children took the sweets' actual location (bottom of the cupboard) into account or not, the boy could reach them and took them to his room.

Experiment 1 (see Fig. 2) was designed so that (a) we could ask both a false belief question and a counterfactual question about the same situation and (b) BRC errors would be detectible in the discriminating conditions for both questions. Then we could check whether changing the test from a discriminating condition to a nondiscriminating condition would change error rates and whether the kinds of errors children made on the two questions pattern together.

In summary, the traditional first-order false belief task is typically mastered around 4 years of age. Children who fail on this task commit the reality error, but the task does not discriminate between BRC and MRC answers. The counterfactual literature recommends such a distinction because children aged 5 years and well beyond typically resist the pull of reality but predominantly base their answers to counterfactual questions on BRC rather than on MRC. Ample evidence of a close link between counterfactual and false belief questions over and above executive functioning and verbal skills (see Fig. 1) prescribes that BRC errors may be possible in false belief tasks and that the use of nondiscriminating conditions in traditional false belief tasks in the past may have been subject to false positive answers.

Experiment 1 tested the hypothesis that false belief reasoning depends on counterfactual reasoning. We predicted that if counterfactual reasoning plays an important role in false belief reasoning, and if participants are old enough to avoid reality errors, then (1) the predominant errors on the counterfactual and false belief questions would be BRC errors, (2) the type of answers to the two questions would correlate substantially, and (3) there would be fewer correct answers to both kinds of question under *discriminating conditions* where BRC and MRC yield different answers than under *nondiscriminating conditions* where BRC and MRC yield the same answer.

Teleology-in-perspective would predict that children should give the same answer to the belief question as to the counterfactual question (Perner & Roessler, 2010). Once it has been established that the mother mistakenly thinks that the little girl had been looking for the sweets, further teleological reasoning is assumed to be carried out counterfactually. Because the mother believes that *the girl had come to look for the sweets on the high shelf,* one can reason counterfactually, "*If the girl had come to look for the sweets on the high shelf,* then [for the MRC reasoner] the sweets would have stayed on that shelf." Hence, the mother would have good reason to think that the sweets are still on the shelf. And that is what MRC reasoners should give for an answer. A BRC reasoner, however, would come to the counterfactual conclusion that the sweets would be in the girl's room. Accordingly, for them the mother would have good reason to think that the sweets were in the girl's room and therefore would give that answer to the belief question.

## Method

### Participants

We recruited 80 healthy participants: 20 7- and 8-year-olds ($M_{age}$ = 8;1 [years;months], *SD* = 0;7, range = 7;0–8;10; 8 girls), 20 9- to 11-year-olds ($M_{age}$ = 10;1, *SD* = 0;5, range = 9;3–11;0; 10 girls), 20 12- to 14-year-olds ($M_{age}$ = 13;6, *SD* = 0;10, range = 12;4–14;10; 8 girls), and 20 adults ($M_{age}$ = 26;3, *SD* = 2;8, range = 20;9–32;4; 17 women). Children were tested in after-school clubs of a medium-sized city and from rural areas. Parents gave written consent. Adults were recruited through opportunity sampling. All participants were German speaking and middle social class of Caucasian origin. The study was approved by the magistrate of Salzburg, Austria.

### Experimental setup

We used two wooden models on 42 × 30-cm platforms. The *sweets story* model had a cupboard with a high shelf in the center and bottom, two female dolls (tall mother and short daughter), and

one male doll (twice as tall as the short female doll). The short daughter could only reach the sweets on the bottom of the cupboard, whereas the male doll could also reach sweets on the high shelf. Each child's room was represented with a table and a photo of each character. The *squirrel story* model had a stone hut and a basket outside. A farmer put nuts either into the hut or basket. A dwarf could get nuts from either location, but a squirrel could only get nuts from the basket, not from the hut.

*Experimental design*

Children were tested in a 20-min session. They saw two discriminating trials and two nondiscriminating trials: one of each per story model. For each trial, several control questions and two test questions were asked: a false belief question (e.g., "Where does mum think the sweets are now?") and a counterfactual question (e.g., "If the little girl had come instead of the tall boy, where would the sweets be now?"). Participants answered two counterfactual questions and two false belief questions per condition. The sequence of the stories (sweets and squirrel), the sequence of the conditions (discriminating and nondiscriminating), and the sequence of the test questions (counterfactual and false belief) were counterbalanced across participants. Orders 1 and 2 received the discriminating conditions first, whereas Orders 3 and 4 received the nondiscriminating conditions first. Orders 1 and 4 received the sweets story first, whereas Orders 2 and 3 received the squirrel story first. Finally, Orders 1 and 3 received false belief questions first, whereas Orders 2 and 4 received counterfactual questions first. Each order contained 5 children per age group.

*Experimental procedure*

In the training phase of the sweets story, children learned that the girl cannot reach the high shelf but can reach the sweets on the bottom of the cupboard. When she finds sweets on the bottom, she brings them into her room; sweets on the high shelf are left in the cupboard. The boy can reach both locations, and when he finds sweets in either location, he brings them to his room (see Fig. 2, training phase). Seven control questions were asked: (1) Which is the boy's room? (2) Which is the girl's room? (3) Which location can the boy reach? (4) Where does he take the sweets? (5) Which location can the girl reach? (6) Where does she take the sweets? (7) Why can't the girl reach the high shelf?

In the test phase, *discriminating condition* (see Fig. 2, test phase, right side), the mother put sweets on the high shelf and left for the living room, and then the boy came to get them. The boy ducked down to appear to be small when entering and leaving the kitchen to make mum believe that his sister took the sweets. Children were told, "Today mum put some sweets on the high shelf. The bottom of the cupboard is empty." Three control questions were asked: (8) Where did mum put the sweets? (9) Would the boy be able to get the sweets? (10) Would the girl be able to get the sweets? Then the boy came looking for sweets. He ducked when entering the kitchen (so that mum would believe his sister came), took the sweets from the shelf, and left the kitchen, ducking again and then disappearing into his room. From the corner of her eye, it appears to mum that someone small had left the kitchen. Children were asked the following questions: (11) Who does mum think has been in the kitchen searching for sweets? This was intended as a standard false belief question. "The girl" was the expected answer. (12) Who was really in the kitchen? "The boy" was the expected answer. At the point when the boy left the sweets on the table in his room and mum had turned away, children were asked two further control questions and two test questions (order of false belief questions and counterfactual question counterbalanced):

False belief: "Where does mum think the sweets are now?"
Control 13: "Where did mum put the sweets in the beginning?"
Control 14: "Where are the sweets really?"
Counterfactual: "If the little girl had come instead of the tall boy, where would the sweets be now?"

Because the little girl cannot reach the high shelf, the correct answer to the counterfactual question is "high shelf." Because mum thinks the little girl came, the correct answer to the false belief question is also "high shelf." The BRC answer to the counterfactual question is "girl's room" (Rafetseder et al., 2010). If children imagine a scenario where mum put the sweets on the bottom, then given that the girl came, they must be in her room now. We were interested in whether children also answer the false belief question with "girl's room," whether we can predict BRC errors on false belief questions

from BRC errors on counterfactual questions, and whether we see more errors on false belief questions in the discriminating condition than in the nondiscriminating condition. Note that reality errors ("boy's room") are still possible and that "bottom of the cupboard" is also a potential error. Wrong answers were not corrected.

After answering these questions, children saw a second episode (e.g., the nondiscriminating condition) of that story. In the nondiscriminating condition (see Fig. 2, test phase, left side), mum put sweets on the bottom of the cupboard and left for the living room. Then the girl, now wearing the boy's sweater, came to the kitchen to search for sweets. She wore her brother's sweater to make mum, who only saw her from the corner of her eye, believe that the boy took the sweets. The girl took the sweets to her room, but this time mum falsely believed it was the boy. Children were asked Control Questions 8 to 14 and two test questions: "Where does mum think the sweets are now?" (false belief) and "If the tall boy had come instead of the little girl, where would the sweets be now?" (counterfactual).

Then the procedure was repeated with the squirrel story. In this story, a farmer put nuts either into the stone hut, where the dwarf collected them and took them to the dwarf village (discriminating condition), or into the basked in front of the stone hut, where the squirrel collected them and took them to a nest in a tree (nondiscriminating condition). The farmer mistook the dwarf, who had a red bag for collecting nuts, for the squirrel, who had a red scarf around its neck, and vice versa.

*Results and discussion*

As predicted, (1) BRC errors predominated on both false belief questions and counterfactual questions and appeared at similar rates on both, (2) the answers children gave to the two questions correlated substantially, and (3) for both kinds of question, correct answers were more common in nondiscriminating conditions than in discriminating conditions.

*Control questions*

Only 3 children made errors on control questions. Each one gave exactly one wrong answer to Control Question 13 ($n = 1$) or 14 ($n = 2$). No mistakes occurred on the standard false belief question (Who does mum think has been in the kitchen searching for sweets?). The stories were equally difficult—within conditions—for counterfactual questions and false belief questions ($p > .05$), so data were collapsed.

*Test questions*

Due to ceiling performance in the nondiscriminating condition, variances differed strongly in the two conditions. Hence, separate mixed factorial repeated-measures analyses of covariance (ANCOVAs), 2 (Question Type: counterfactual or false belief) × 4 (Age Group) × 4 (Order: 1, 2, 3, or 4), were conducted on the number of correct responses to test questions, with question type as a within-participants factor, age group as between-participants factor, and order as a covariate. In the discriminating condition, significant main effects were obtained for age group, $F(3, 75) = 11.96$, $p < .001$, $\eta_p^2 = .32$, and for order, $F(1, 75) = 9.25$, $p = .003$, $\eta_p^2 = .11$. We did not find a main effect of question type, $F(1, 75) = 3.05$, $p = .085$, $\eta_p^2 = .04$, or a significant interaction (lowest $p = .14$).

In terms of the main effect of age, Bonferroni-adjusted post hoc analysis revealed a significant difference in performance of the 8-year-olds compared with all the other age groups (10-year-olds: $M_{diff} = 0.53$, 95% confidence interval (CI) [0.01, 1.04], $p = .046$; 13-year-olds: $M_{diff} = 0.75$, 95% CI [0.23, 1.27], $p = .001$; adults: $M_{diff} = 1.13$, 95% CI [0.61, 1.64], $p < .001$). In addition, the 10-year-olds differed significantly from the adults ($M_{diff} = 0.60$, 95% CI [0.08, 1.12], $p = .015$), but not from the 13-year-olds ($M_{diff} = 0.23$, 95% CI [−0.29, 0.74], $p > .99$). Finally, the 13-year-olds did not differ from the adults ($M_{diff} = (0.38$, 95% CI [−0.14, 0.89], $p = .319$). Thus, children got progressively better with age until about 12 to 14 years, when they reached adult-level performance.

We also found an effect of order. Bonferroni-adjusted post hoc analysis revealed a significantly low performance in Order 4 (sweets story first, counterfactual question first, nondiscriminating condition first) compared with every other order (Order 1: $M_{diff} = 0.63$, 95% CI [0.11, 1.14], $p = .01$; Order 2: $M_{diff} = 0.68$, 95% CI [0.16, 1.19], $p = .004$; Order 3: $M_{diff} = 0.70$, 95% CI [0.18, 1.22], $p = .003$). All other orders did not differ significantly from each other. This effect is very specific in that it indicates that

each factor individually (Order 1: sweets story first; Order 2: counterfactual first; Order 3: nondiscriminating condition first) did not lower children's performance, only the accumulated combination in Order 4.

The same ANCOVA for the nondiscriminating condition showed neither a main effect of question type, $F(1, 75) = 0.21$, $p = .65$, $\eta_p^2 = .03$, nor a main effect of age group, $F(3, 75) = 0.52$, $p = .671$, $\eta_p^2 = .02$, or order, $F(1, 75) = 0.30$, $p = .59$, $\eta_p^2 = .004$. No interactions were significant (lowest $p = .17$).

Table 1 shows a clear interaction between condition and age that was not captured by the separate analyses for each condition. So, we conducted a one-way analysis of variance (ANOVA) on the difference score between the discriminating and nondiscriminating conditions with age group as between-participants factor. This showed a significant main effect of age, $F(3, 76) = 6.50$, $p = .001$, $\eta_p^2 = .21$. Bonferroni-adjusted post hoc analysis revealed that the 8-year-olds differed significantly from the 13-year-olds ($M_{diff} = 1.50$, 95% CI [0.18, 2.82], $p = .017$) and from the adults ($M_{diff} = 2.05$, 95% CI [0.73, 3.37], $p < .001$), indicating that the difference between conditions decreased with age. All children gave more correct answers in the nondiscriminating condition than in the discriminating condition, lowest $t(19) = 2.59$, $p = .018$.

Moreover, the answers to the counterfactual and false belief questions were significantly correlated ($r = .56$, $n = 80$, $p < .001$; when controlling for age: $r = .47$, $p < .001$). This was our second predicted finding. This correlation was significant even when assessed in a more conservative fashion across stories ($\Phi = .34$, $n = 160$, $p < .001$), which speaks against the possibility that children just tended to repeat their answer to the first question on the second question.

*Response types (error analysis)*

Because responses in the nondiscriminating condition were mostly correct (Table 1), we present the error analysis only for the discriminating condition. Fig. 3 shows children's answers to false belief and counterfactual questions broken down by error type. Just visually inspecting the graph, one can see that MRC responses steadily increase with age, whereas BRC errors decrease. As expected within this age range, other types of errors were close to zero.

The marginal sums of Table 2 show that BRC errors predominated and were committed consistently on both false belief questions (21 children) as often as on both counterfactual questions (15 children) (McNemar's $p = .21$). This was our first predicted finding. The table (shaded cells) also shows the consistency of participants' answers to counterfactual and false belief questions. In sum, 39 (48.8%) of all participants gave the same type of answer to all four test questions (two counterfactual questions and two belief questions), and 15 more (18.8%) gave the same answer to three of the four test questions. We found a strong contingency between strategies used to answer the counterfactual question and those used to answer the belief question ($\Phi = .49$, $n = 160$, $p < .001$), which confirms our second prediction even when assessed in a more conservative fashion across stories ($\Phi = .42$, $n = 160$, $p < .001$).

The data fully support the three predictions we drew from the assumption that counterfactual reasoning is necessary for false belief reasoning: (1) BRC errors predominate and appear at similar rates, (2) the type of answers to the two questions correlate substantially, and (3) correct answers are much lower for the discriminating condition than for the nondiscriminating condition.

One might object that our task was excessively complex. We added complexity to make the BRC error possible. Having shown the basic pattern we predicted, we replicated the core findings with a simpler design in Experiment 2.

## Experiment 2

We made five changes to simplify and improve the design in light of the findings from Experiment 1. First, (1) we eliminated the nondiscriminating condition because performance was at ceiling. This enabled us to (2) make the source of mum's false belief more natural and (3) simplify the structure of the story by eliminating one of the characters. In addition, (4) we adjusted the false belief question so that it explicitly mentioned mum's missing information; this improved the match between the false belief question and the counterfactual question (Krzyżanowska, 2013). Finally, (5) we made the event

**Table 1**
Mean numbers of correct responses (and standard deviations) on the two test questions of each condition (discriminating and nondiscriminating) split for age group in Experiment 1.

| | | Age group | | | |
|---|---|---|---|---|---|
| Condition | Question | 7–8 years | 9–11 years | 12–14 years | Adults |
| Discriminating | FB | 0.40 (0.75) | 0.95 (0.83) | 1.40 (0.82) | 1.70 (0.57) |
| | CF | 0.85 (0.88) | 1.35 (0.81) | 1.35 (0.75) | 1.80 (0.52) |
| Nondiscriminating | FB | 1.85 (0.37) | 1.95 (0.22) | 1.90 (0.31) | 1.95 (0.22) |
| | CF | 1.80 (0.62) | 1.85 (0.37) | 1.75 (0.44) | 1.90 (0.31) |

*Note.* FB, false belief; CF, counterfactual.

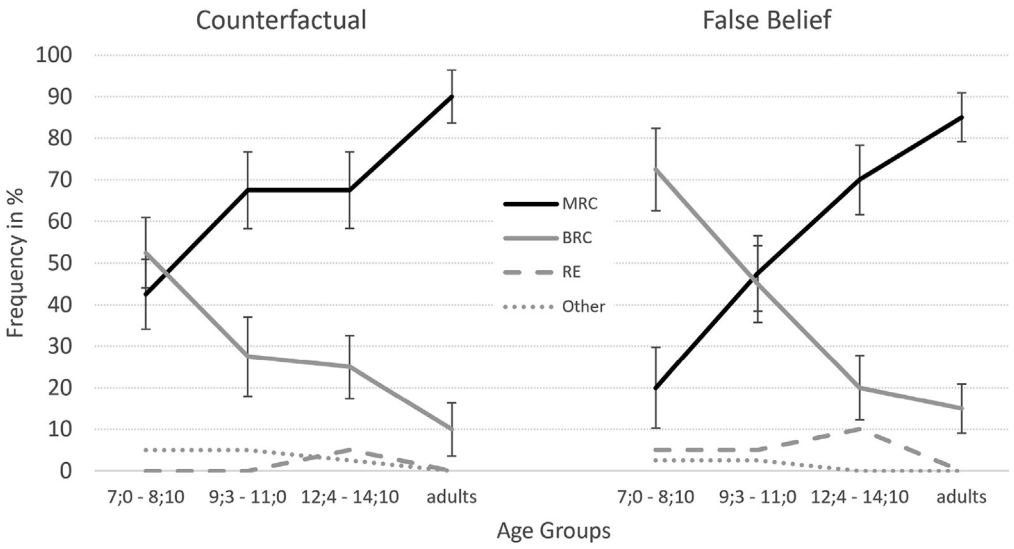


**Fig. 3.** Frequency of answers (%) per age group and question type in Experiment 1. Other, indicating the other location.

**Table 2**
Consistency of answers to counterfactual and false belief questions across both stories of the discriminating condition in Experiment 1.

| Counterfactual questions | False belief question | | | | **Total Counterfactual** |
|---|---|---|---|---|---|
| | 2 × MRC | 2 × BRC | MRC / BRC | Other | |
| 2 × MRC | 29 | 5 | 8 | 2 | **44** |
| 2 × BRC | 3 | 10 | - | 2 | **15** |
| MRC / BRC | 2 | 5 | 5 | 3 | **15** |
| Other | 2 | 1 | 2 | 1 | **6** |
| **Total False Belief** | **36** | **21** | **15** | **8** | **80** |

*Note.* MRC, mature reasoning with counterfactuals; BRC, basic reasoning with counterfactuals. Other combinations include reality errors and errors not further defined (see Fig. 2, answer options). The shaded cells show the consistency of participants' answers to counterfactual and false belief questions.

that engenders mum's false belief into the kind of exceptional circumstance that triggers counterfactual thoughts more easily (Kahneman & Miller, 1986, p. 143) in order to help children initiate counterfactual reasoning.

Because our simplifications may improve performance, we also included 3- to 6-year-old participants.

## Method

### Participants

We piloted the procedure with 6 English-speaking adults ($M_{age}$ = 37.33 years, *SD* = 17.78) to ensure that all questions were clear. Subsequently, 86 children and adolescents participated. One 4-year-old boy was excluded because he did not respond to the main test questions. Ages in the final sample ranged from 3;11 to 12;9 ($M_{age}$ = 7;3, *SD* = 2;4; 39 girls). They were recruited from two primary schools and one playgroup in the United Kingdom ($M_{age}$ = 5;10, *SD* = 1;2; *n* = 55) (first language English) and from one primary and one secondary modern school in Austria ($M_{age}$ = 9;11, *SD* = 1;8; *n* = 30) (first language German). All participants were middle social class of Caucasian origin.

We split the sample into three groups: 32 3- to 5-year-olds ($M_{age}$ = 5;0, range = 3;11–5;11; 12 girls), 36 6- to 8-year-olds ($M_{age}$ = 7;6, range = 6;0–8;11; 19 girls), and 17 9- to 12-year-olds ($M_{age}$ = 11;1; range = 9;3–12;9; 8 girls). Parents consented to participation in writing. The experiment was approved by the psychology ethics committee of the University of Stirling.

### Experimental design

Each participant was tested in one 10-min session in a quiet area away from other children. The participant was given both the sweets and squirrel stories. Each participant answered two false belief questions and two counterfactual questions. Order of test questions and stories was counterbalanced within participants.

### Experimental procedure

In this version of the sweets story, there was only one child and only the shelf was used. Mum puts sweets on the shelf every day, and when the boy finds sweets, he takes them to his room. After learning these rules, 3- to 8-year-olds were asked two control questions—(C1) "Where does mum put the sweets every day?" and (C2) "What happens when the boy comes in search for the sweets?"—to check for memory lapses. Then they saw mum put sweets on the shelf. After she left, the sweets slid off the shelf and the dog took them to the garden. At this point, mum falsely believes that the sweets are on the shelf, and the source of that belief is more natural than in Experiment 1. When the boy comes home, he sadly finds no sweets on the shelf. Mum sees him leave the kitchen and disappear into his room. Children were asked the following (test questions counterbalanced):

False belief: "Mum did not see the sweets fall off the shelf. Where does mum think the sweets are?"
Control 3: "Where are the sweets really?"
Counterfactual: "If the sweets had not fallen off the shelf, where would the sweets be?"
Control 4: "Where did mum put the sweets?" (only 3- to 8-year-olds were asked this question).

In the squirrel story, a farmer put some nuts into a basket on a shelf in a hut. Usually a squirrel squeezes through a hole in the hut and takes the nuts to the tree. After the farmer left, the basket fell off the shelf and the nuts rolled out of the hut. A boy came by, found the nuts, and took them home. Later the squirrel came and could not find any nuts. The farmer saw the squirrel leave the hut and disappear behind the tree trunk. Children were then asked a false belief question ("The farmer did not see the nuts fall off the shelf. Where does the farmer think the nuts are?") and a counterfactual question ("If the nuts had not fallen off the shelf, where would the nuts be?").

Note that we did not include the word 'now' in the false belief question (as in Experiment 1) so that the youngest children would not be tempted to indicate the real location of the sweets. Also note that reality errors ("in the garden" and "at home") differ from BRC errors ("on the shelf") and from the MRC answers ("in the boy's room" and "at the tree"), where BRC errors are the result of violations of the nearest possible world constraint.

*Results and discussion*

*Pilot study and control questions*

In the pilot study, adults answered all questions correctly. In the actual experiment, children answered 98% of control questions correctly. Mistakes were evenly distributed and happened mainly in the 3- to 5-year-old age group. Keeping these children in the sample did not affect the results. The two stories were of comparable difficulty for both test questions ($p > .05$); hence, data were collapsed. Question order had no effect on the number of correct answers either to counterfactual questions (Fisher's exact; sweets story: $p = .83$; squirrel story: $p = .13$) or to false belief questions (sweets story: $p = .66$; squirrel story: $p = .052$) and is not considered further.

*Test question*

Fig. 4 shows that children gave similar numbers of correct answers to counterfactual questions as to false belief questions. A mixed factorial repeated-measures ANOVA, 3 (Age Group) × 2 (Question Type), on the number of correct responses to test questions with age group as between-participants factor showed a significant main effect of age group, $F(2, 82) = 41.26$, $p < .001$, $\eta_p^2 = .50$, but not of question type, $F(1, 82) = 2.06$, $p = .16$, $\eta_p^2 = .02$. The interaction was not significant, $F(2, 82) = 2.21$, $p = .12$, $\eta_p^2 = .05$. Bonferroni-adjusted post hoc analysis revealed that the 3- to 5-year-olds differed significantly from the 6- to 8-year-olds ($M_{diff} = 1.89$, 95% CI [1.18, 2.63], $p < .001$) and from the 9- to 12-year-olds ($M_{diff} = 3.24$, 95% CI [2.32, 4.15], $p < .001$), and that the 6- to 8-year-olds differed significantly from the 9- to 12-year-olds ($M_{diff} = 1.35$, 95% CI [0.45, 2.24], $p = .001$), indicating a clear developmental progression.

Like in Experiment 1, performance on both test questions was significantly and very highly correlated ($r = .73$, $n = 85$, $p < .001$; when controlling for age, $r = .56$, $p < .001$) even across stories ($\Phi = .53$, $n = 170$, $p < .001$). This was our second predicted finding.

*Response types (error analysis)*

The marginal sums of Table 3 show that the BRC error was committed consistently on both counterfactual questions and on both false belief questions. This confirmed our first prediction. The shaded cells show the consistency of participants' answers to counterfactual and false belief questions. In

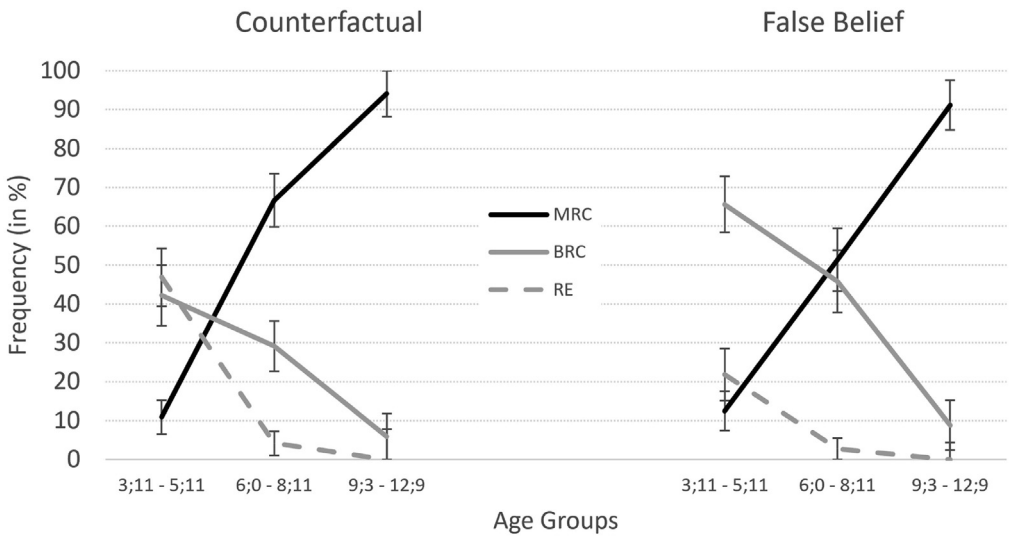


**Fig. 4.** Frequency of answers (%) per age group and question type in Experiment 2. Error bars represent standard errors. MRC, mature reasoning with counterfactuals; BRC, basic reasoning with counterfactuals; RE, reality error.

**Table 3**
Consistency of answers to counterfactual and false belief questions across both stories in Experiment 2.

| Counterfactual questions | False belief question | | | | | | Total Counter-factual |
|---|---|---|---|---|---|---|---|
| | 2 × MRC | 2 × BRC | MRC / BRC | 2 × RE | MRC / RE | BRC / RE | |
| 2 × MRC | 29 | 5 | 3 | - | - | - | **37** |
| 2 × BRC | - | 12 | 1 | 2 | - | 2 | **17** |
| MRC / BRC | 4 | 4 | 1 | - | - | - | **9** |
| 2 × RE | 1 | 4 | 2 | 3 | - | 1 | **11** |
| MRC / RE | - | 2 | 1 | - | - | 1 | **4** |
| BRC / RE | - | 6 | - | 1 | - | - | **7** |
| **Total False Belief** | **34** | **33** | **8** | **6** | **0** | **4** | **85** |

*Note.* MRC, mature reasoning with counterfactuals; BRC, basic reasoning with counterfactuals; RE, reality error. The shaded cells show the consistency of participants' answers to counterfactual and false belief questions.

sum, 44 participants (51.8%) gave the same type of answer to all four test questions. The contingency between reasoning strategies for answering the counterfactual question and the belief question was high ($\Phi$ = .57, $n$ = 170, $p$ < .001; across stories, $\Phi$ = .54, $n$ = 170, $p$ < .001), which again confirmed our second prediction. Because only discriminating conditions were used in this experiment, we could not test our third prediction.

*Comparing Experiments 1 and 2*

In Experiment 2, we reduced task complexity and changed the false belief question to explicitly mention what mum has not seen. We conducted a one-way ANCOVA on the number of correctly answered false belief questions with experiment as a between-participants factor and age as a covariate (adult samples were not included). This showed a significant main effect, $F(1, 142) = 17.05$, $p < .001$, $\eta_p^2 = .11$, with the number of correct answers increasing from Experiment 1 ($M$ = 0.52, $SE$ = 0.11) to Experiment 2 ($M$ = 1.17, $SE$ = 0.09). The same analysis for counterfactual questions also returned a significant main effect, $F(1, 142) = 7.85$, $p = .006$, $\eta_p^2 = .05$ (Experiment 1: $M$ = 0.84, $SE$ = 0.11; Experiment 2: $M$ = 1.27, $SE$ = 0.09). Thus, the changes implemented in Experiment 2 improved children's performance on both test questions and reduced BRC errors for counterfactual questions, $F(1, 142) = 7.15$, $p = .008$, $\eta_p^2 = .05$) as well as for false belief questions, $F(1, 142) = 12.02$, $p = .001$, $\eta_p^2 = .08$.

The test questions were significantly easier to answer in Experiment 2 than in Experiment 1. Yet the predictions we made from the assumption that counterfactual reasoning is a prerequisite for false belief reasoning were again confirmed: Answers to the two questions contain predominantly BRC errors, and the type of answers to the two questions correlate substantially.

## General discussion

In existing studies, a relationship between counterfactual reasoning and belief reasoning is correlational. Our experiments introduced a manipulation; we varied the difficulty of the reasoning required to answer the counterfactual question correctly and checked whether children's answers to false belief questions varied accordingly. We found that (1) BRC errors predominated and occurred at least as often on the false belief questions as on the counterfactual questions, (2) answer types were strongly correlated, and (3) much fewer correct responses appeared in discriminating conditions than in nondiscriminating conditions. We also found (4) more correct responses in the discriminating condition of Experiment 2 than in Experiment 1.

We now discuss three broad strategies for explaining these findings. We assess (1) domain-general accounts, (2) general theory-of-mind frameworks, and (3) domain-specific accounts. We conclude with our own general theory-of-mind framework—teleology-in-perspective with sensitivity to the nearest possible world constraint.

We start with two domain-general accounts: Could our findings reflect changes in processing capacity or inhibition?

### *Changes in processing capacity*

First, consider processing capacity. Reducing processing demands in standard false belief tasks improves children's performance (Bartsch, 1996; Chandler, Fritz, & Hala, 1989; Friedman & Leslie, 2005; Hansen, 2010; Lewis & Osborne, 1990; Mitchell & Lacohée, 1991; Rubio-Fernández & Geurts, 2013; Setoh, Scott, & Baillargeon, 2016; Westra, 2017). Thus, processing demands are likely playing some role in explaining the relation between false belief and counterfactual reasoning. Working memory can partially explain this association, but the relevant findings pertain only to studies where children could commit the reality error but not the BRC error (Drayton, Turley-Ames, & Guajardo, 2011; Guajardo, Parker, & Turley-Ames, 2009). This is plausible for the standard false belief task where working memory overload makes children fall back on the typical prepotent default answer—the reality error. If there is no prepotent default, then working memory overload should result in unsystematic error patterns. Prepotency holds for the reality error in our tasks, but not for the BRC error. The BRC response is not an obvious default; it only becomes prevalent for counterfactual questions as a result of ignoring the nearest possible world constraint. So, if working memory capacity and inhibition are good enough to avoid the reality error but are not good enough to make children obey the nearest possible world constraint (even though they know how to use it), it is unclear why they should produce the BRC error in response to the false belief question unless they approach it by counterfactual reasoning and fail to keep to the nearest possible world constraint.

Alternatively, in the discriminating conditions of Experiment 1, children may simply forget that the sweets were on the high shelf where the girl cannot reach them. In the nondiscriminating condition, forgetting does not matter because the boy can reach the sweets either way. In Experiment 2, children who forget that the boy came looking for sweets and that mum saw him do so will answer "on the shelf" to both questions. Children who do not forget will indicate the boy's room. Yet the better performance in Experiment 2 than in Experiment 1 does not fit. In Experiment 1, the antecedent of the counterfactual question explicitly mentions that someone was searching for sweets, which should help children to remember the sweets' location. In Experiment 2, participants must remember that the boy came looking for sweets, but the counterfactual antecedent gives them no hint that would help them to remember this. Moreover, the finding that children's working memory capacity accounts for little variance in the data on various counterfactual tasks speaks against this speculation (Beck, Riggs, & Gorniak, 2009). Finally, Rafetseder et al. (2010) showed that asking children about the sweets' earlier location (e.g., what if mum had placed the sweets on the high shelf rather than on the bottom?) did not improve performance in a discriminating condition, which would be predicted if children had simply forgotten the sweets' initial location.

### *Changes in inhibition*

Poor inhibition may explain why children make reality errors on false belief tests as well as on counterfactual tasks (Beck et al., 2009). Inhibition of a default response, such as the reality error, is ubiquitous when attributing a belief with false content to an agent. In our tasks, children must inhibit the reality answer as well as the BRC answer. Adding extra inhibitory steps makes false belief tasks harder. For example, tasks that require *double* inhibition—that is, disinhibition of an inhibited location (e.g., avoid full location and then avoid empty location because it is thought to be full)—are more difficult (Leslie & Polizzi, 1998). Extra inhibitory demands could in theory (a) explain the correlation between false belief and counterfactual questions and (b) explain the difference between conditions if questions in the nondiscriminating condition require only "single inhibition."

However, the problem of double inhibition does not apply in our tasks for two reasons. First, there is no need for *meta*inhibition; the inhibited real location (e.g., in the garden with the dog) should never be disinhibited. Second, there is no indication of the need for *serial* inhibition; the predominant error in the discriminating condition is not a reality error. It is not obvious why resisting the BRC error should tax inhibition (the fact that "on the shelf"—the BRC response—was mentioned in both test questions of Experiment 2 did not lead to more BRC errors than in Experiment 1) or why this should be more difficult than inhibiting reality answers.

We now turn to the question of whether our results could be explained within the general frameworks of theory of mind, namely theory theory and simulation theory.

*Theory theory*

According to theory theory, mental states such as beliefs and desires are theoretical entities. Quasi-lawlike generalizations govern these theoretical entities. We use these generalizations to predict other people's beliefs, desires, emotions, and actions. The laws need to cover when mental states are to be attributed and how they generate behavior. How belief attribution is thought to work can be gleaned from this example from Carruthers (2013):

> Consider a false-belief task: a doll that an agent has been playing with is first placed in a blue box and then, while the agent is absent, is moved to a green box. . . . [Note] that the agent is not present when the doll is moved. . . . During the initial sequence the infant infers [that] *the agent thinks: the doll is in the blue box,* relying on the attributional principle, *seeing leads to believing*, or some-such. It then does not update this representation when the doll is moved [because the agent was absent. The representation thus stays the same: *the agent thinks: the doll is in the blue box*]. (p. 160)

Counterfactual reasoning has no role in this account because all mental states attributed are supposed to be the mental states the agent actually has. Neither does counterfactual reasoning become relevant for predicting where the agent will go to get the doll. For this, theory of mind makes use of the practical syllogism:

> These tenets [the main tenets of the "adult theory"] are perhaps best summarized by the "practical syllogism": "If a psychological agent wants event *y* and believes that action *x* will cause event *y*, he will do *x*." Many philosophers have argued that the practical syllogism is the basic explanatory schema of folk psychology. (Gopnik & Meltzoff, 1997, p. 126)

In our example, this means that the agent wants to be where the doll is and believes that the way to achieve this is to go to the blue box where he believes the doll to be. Again, it is not apparent how any counterfactual reasoning could be of help here.

Gopnik and Wellman (2012) made the point that a fully causal theory must incorporate the ability for counterfactual reasoning:

> Theories have distinctive cognitive functions. They . . . allow you to make counterfactual inferences—inferences about . . . what would happen if you decided to intervene on the world and do something new in the future. These inferences about counterfactuals and interventions go beyond simple predictions about what will happen next. (p. 1086)

Because the child's theory of mind is supposed to be a theory, it suggests that there ought to be a place for counterfactual thinking. And there is, but at a different level of causal connections than is required in our experiments. Our questions concerned events in the world ("If the girl had come instead of the boy, where would the sweets be?"), whereas the questions Gopnik and Wellman (2012) have in mind are about the lawful connection of mental states about the world: *If the agent had seen the transfer, where would he go? If the agent hated the doll, what would he do?* This kind of counterfactual reasoning is expedient for intervening in agents' mental processing but evidently addresses a different level of causal connections than our counterfactual questions and is not needed to answer our false belief questions.

*Simulation theory*

Simulation theory proposes that we calculate others' mental states by using our own cognitive apparatus "offline." We adjust the inputs to our system to replicate the input to someone else's cognitive system, let our system run, and attribute those outputs to the person we were simulating. A standard suggestion (Goldman, 2006) is that pretend mode quarantines simulated beliefs from one's real beliefs.

Simulation theory speaks to false belief but not to counterfactuality. It might explain the relationship between counterfactual and false belief questions if pretense is a form of counterfactual reasoning, as suggested by Buchsbaum, Bridgers, Skolnick Weisberg, and Gopnik (2012; see also Weisberg & Gopnik, 2013). Beck (2016) raised issues for this claim, arguing that counterfactuals, unlike pretense, are "closely related to and impact on reality. They . . . are evaluated relative to the events they replace" (p. 254).

But even if pretense is counterfactual reasoning, it is hard to justify why simulating mum's belief in the discriminating condition is more difficult than it is in the nondiscriminating condition, particularly when the mother herself does not reason counterfactually to arrive at her mistaken belief. For example, in Experiment 1 mum puts the sweets on the shelf. Later she sees someone small enter the kitchen. She will conclude that the sweets will stay put because her daughter cannot reach the shelf. Mum does not use counterfactual reasoning here, so it is unclear why a simulator (once the alleged counterfactual step of pretending has been successful) would encounter difficulties attributable to differences in counterfactual reasoning. As with theory theory, it is unclear how simulation theory would justify additional assumptions required to predict that children will commit the same type of error on the counterfactual question and the false belief question.

We now turn to domain-specific accounts of our data.

*Known problems with belief understanding in middle childhood: Second-order belief and inference neglect*

Children struggle with second-order false beliefs long after they pass the original false belief task (Astington, Pelletier, & Homer, 2002; Coull, Leekam, & Bennett, 2006; Perner & Howes, 1992; Perner & Wimmer, 1985; Sullivan, Zaitchik, & Tager-Flusberg, 1994). Participants in our experiments need not monitor beliefs about beliefs. They must understand that mum combines two beliefs (the sweets were on the shelf and her son went into the kitchen) to infer where the sweets are (in the boy's room). So, it seems unlikely that children's problems arise from struggles with second-order beliefs. However, this raises another problem. Sodian and Wimmer (1987) showed that children do not realize that others make inferences for several years after they fail the original false belief task. Perhaps the children fail our task because they do not realize that others draw inferences. This suggestion is interesting but incomplete, in particular because it does not explain why this should affect the difficulty of the counterfactual question to the same degree. Not realizing that others draw inferences would make our task difficult. But why should it so systematically yield BRC errors? In the discriminating condition of Experiment 1, mum believes that the sweets are on the shelf and sees someone small entering the kitchen. Mum does not need to draw an inference to arrive at the belief that the sweets are on the shelf; she already believes that. She needs to avoid changing her existing beliefs. It is not obvious how this would be explained in terms of inference neglect.

*Adaptive modeling*

According to adaptive modeling (Peterson & Riggs, 1999), we store our information about the world in a database that can be queried to answer questions. Modeling someone else's mental states requires manipulating this database through a four-step process: (1) identifying differences between one's own database and the database of the person being modelled, (2) temporarily implementing those differences in one's own database, (3) using the adapted database to answer questions, and (4) attributing the output to the individual whose mental states were modeled. Answering counterfactual questions requires a similar three-step process: (1) temporarily implementing the antecedent in the database,

(2) using the adapted database to answer the question in the consequent, and (3) treating the result as the answer to the counterfactual question.

Table 4 applies Peterson and Riggs's (1999) analysis to our false belief test of the nondiscriminating condition (left) and the discriminating condition (right) of Experiment 1.

In the nondiscriminating condition, mum put sweets beneath the shelf. The girl, wearing the boy's sweater, came and took them to her room. These facts plus accompanying norms are stored in our participant's database. To calculate mum's false belief, the participant needs to (1) establish information that mum lacks (i.e., that it was the girl who was wearing the boy's sweater), (2) remove that information from the database (grayed out in Table 4), (3) run the query "Where are the sweets?" on the modified database, and (4) attribute the result—that the sweets are in the boy's room—to mum. Answering counterfactual questions in this condition (if the tall boy had come instead of the little girl, where would the sweets be?) requires a similar process: (1) assuming that the boy came, not the girl, which requires graying out the information that the person in the boy's sweater was the girl, (2) asking the question posed by the consequent (where are the sweets now?), and (3) treating the result—boy's room—as the answer to the counterfactual question. Although the steps of the procedure are different in the false belief case and the counterfactual case, the database manipulations are identical. Consequently, adaptive modeling predicts correlated performance on counterfactual questions and false belief questions because both rely on a common core set of abilities—the ability to appropriately modify an existing database and run queries on the modified database. Adaptive modeling predicts the same correlation in the discriminating condition (Table 4, right).

However, adaptive modeling does not predict the difference found between the two conditions. The database modifications are similar in both conditions. Thus, children who answer correctly in the nondiscriminating condition should answer correctly in the discriminating condition. Any explanation for why the modifications are difficult in one condition would seem to apply equally well in the other condition, and so adaptive modeling does not predict the difference observed between conditions. This does not rule out common-cause explanations of the correlation, but it speaks against Peterson and Riggs's (1999) explanation as it currently stands. Adaptive modeling could be made more precise so as to make appropriate predictions, and our findings provide guidance for how to develop the theory. In its current form, adaptive modeling does not fully explain why children find the nondiscriminating condition to be easier than the discriminating condition.

### *Teleology-in-perspective*

Like theory theory and simulation theory, teleology is a general framework of what constitutes our understanding of the mind. Unlike these other two frameworks, it explicitly employs counterfactual reasoning to account for differences of perspective, as in false belief reasoning. It is based on the intuition that an action is intentional if good reasons speak in its favor (Anscombe, 1957; Scanlon, 1998). For teleology, these reasons are objective facts that justify the action (Perner & Roessler, 2010). Facts speak in favor of an action only if its goal is something (in a minimal sense) good, desirable, or worth pursuing,[3] and therefore the instrumental actions that achieve this goal become worth carrying out. Without the goal, the actions would not make sense.

We can also use teleological reasoning to model another person's reasoning from that person's perspective. To do so, we ask ourselves counterfactual questions that specify the relevant perspective differences. If one wants to model Max's perspective in the original false belief task, one needs to ask, "If the chocolate hadn't been moved out of the drawer, where would one have objective reason to think the chocolate was?" More generally, one asks, "What objective reason would there be to believe proposition *p* if that person's beliefs were objective facts?" Teleological reasoning that captures alternative perspectives using counterfactuals is called *teleology-in-perspective.*

How does this explain our findings? Take the counterfactual questions first. When asked, "If the little girl had come instead of the tall boy, where would the sweets be now?", people look for a tele-

[3] The requirement that the goal of an action needs to be seen as something good or else nothing can speak in favor of the action is in the spirit of Aristotle's teleology (Charles, 2012) and makes it different from the teleology used in developmental psychology by Csibra and Gergely (1998).

**Table 4**
Adaptive modeling applied to Experiment 1.

| | non-discriminating condition | discriminating condition |
|---|---|---|
| Query | Where are the sweets? | Where are the sweets? |
| Facts | Sweets were on bottom, person in boy's sweater came, person in boy's sweater was the girl | Sweets were on high shelf, small looking person came, small looking person was the boy |
| Norms | Boy usually wears boy's sweater, girl does not usually wear boy's sweater, boy takes sweets to his room, girl takes sweets to her room | Girl looks small, boy looks tall, boy takes sweets to his room, girl cannot reach high shelf |
| Answer | Boy's room | On high shelf |

*Note.* Database of a reasoner with full information. Grayed out information is not known by the agent modeled (e.g., mum).

ologically satisfying answer. Because the point of the girl searching for sweets is to take them to her room, "girl's room" is satisfying. Children using BRC are not bound by any causal constraints of the actual event and can freely opt for this answer, committing the BRC error. Participants who use MRC, however, will realize that the girl's intentions would be thwarted by physical constraints and that the sweets would remain on the high shelf—the correct answer.

In Experiment 2, the test question "If the sweets had not fallen off the shelf, where would the sweets be?" has two teleologically sensible answers. One is "on the shelf" because the mother put the sweets there for them to stay there and not to fall off. The other is "in the boy's room" in case some children reason that the mother put the sweets there to be picked up by the boy. So, children using BRC might commit the BRC error or give the correct answer, which explains why children achieved more correct answers in Experiment 2 than in Experiment 1 (fourth finding).

Consider the false belief question, "Where does mum think the sweets are?" The teleologist assumes that mum holds the beliefs that she has objective reasons to hold. If the sweets are in the boy's room, then mum has objective reasons to believe that that is where they are. This yields the reality error. Children who can take the mother's perspective (teleology-in-perspective) pose themselves a counterfactual question. In the discriminating condition of Experiment 1, where the boy came but the mother thought it was the girl, that question is, "If the little girl had come instead of the tall boy, where would there be objective reason to believe that the sweets are?" So, the counterfactual question and the false belief question will get the same answer. This tells us why BRC errors are the predominant errors to the false belief questions (first finding) and explains the good correlation of answers given to the two questions (second finding).

In nondiscriminating conditions, poor adherence to the nearest possible world constraint does not hinder children's performance. For example, in Experiment 1, the girl takes sweets to her room and participants are asked, "If not the little girl but the tall boy had come, where would the sweets be?" In this condition, participants do not need to keep track of which shelf mum put the sweets on because the boy can reach them either way. So, BRC yields the correct answer. This explains why children gave more correct answers in this condition than in the discriminating condition (third finding). All told, combining teleology-in-perspective with the nearest possible world constraint predicts the range of behaviors observed in our experiments to an impressive degree.

## *Conclusion*

Our results shed light on the latent processes of our "folk psychology," that is, when and how we draw inferences about other people's behavior—their mental states, thoughts, intentions, beliefs, and

desires. We propose that early reasoning is simply teleological; people aim to do what brings about good results. More mature reasoning about others requires that we reason teleologically but from their perspective. This requires treating perspective differences, such as others' false beliefs, as counterfactual suppositions and reasoning from those suppositions. But children at different ages use different strategies; younger children can only use BRC, whereas older children and adults are able to use MRC. This contrasts with standard accounts in theory of mind such as theory theory and simulation theory, where counterfactual reasoning does not play a role, and so the distinction between BRC and MRC cannot be easily put to explanatory work. Our approach accommodates this distinction and makes predictions about the effect of manipulating the use of these types of reasoning. We confirmed those predictions in the experiments reported here.

## References


Anscombe, G. E. M. (1957). *Intention*. Oxford, UK: Blackwell.

Astington, J. W., Pelletier, J., & Homer, B. (2002). Theory of mind and epistemological development: The relation between children's second-order false-belief understanding and their ability to reason about evidence. *New Ideas in Psychology, 20*, 131–144.

Bartsch, K. (1996). Between desires and beliefs: Young children's action predictions. *Child Development, 67*, 1671–1685.

Beck, S. R. (2016). Why what is counterfactual really matters: A response to Weisberg and Gopnik (2013). *Cognitive Science, 40*, 253–256.

Beck, S. R., Riggs, K. J., & Gorniak, S. L. (2009). Relating developments in children's counterfactual thinking and executive functions. *Thinking & Reasoning, 15*, 337–354.

Buchsbaum, D., Bridgers, S., Skolnick Weisberg, D., & Gopnik, A. (2012). The power of possibility: Causal learning, counterfactual reasoning, and pretend play. *Philosophical Transactions of the Royal Society B: Biological Sciences, 367*, 2202–2212.

Carruthers, P. (2013). Mindreading in infancy. *Mind & Language, 28*, 141–172.

Chandler, M., Fritz, A. S., & Hala, S. (1989). Small-scale deceit: Deception as a marker of two-, three-, and four-year-olds' early theories of mind. *Child Development, 60*, 1263–1277.

Charles, D. (2012). Teleological causation. In C. Shields (Ed.), *The Oxford handbook of Aristotle* (pp. 227–266). Oxford, UK: Oxford University Press.

Clements, W. A., & Perner, J. (1994). Implicit understanding of belief. *Cognitive Development, 9*, 377–395.

Coull, G. J., Leekam, S. R., & Bennett, M. (2006). Simplifying second-order belief attribution: What facilitates children's performance on measures of conceptual understanding? *Social Development, 15*, 548–563.

Csibra, G., & Gergely, G. (1998). The teleological origins of mentalistic action explanations: A developmental hypothesis. *Developmental Science, 1*, 255–259.

Drayton, S., Turley-Ames, K. J., & Guajardo, N. R. (2011). Counterfactual thinking and false belief: The role of executive function. *Journal of Experimental Child Psychology, 108*, 532–548.

Eddy, C. M., Beck, S. R., Mitchell, I. J., Praamstra, P., & Pall, H. S. (2013). Theory of mind deficits in Parkinson's disease: A product of executive dysfunction? *Neuropsychology, 27*, 37–47.

Edgington, D. (2004). Counterfactuals and the benefit of hindsight. In P. Dowe & P. Noordhof (Eds.), *Cause and chance: Causation in an indeterministic world* (pp. 12–27). Abingdon, UK: Routledge.

Edgington, D. (2008). Counterfactuals. In *Proceedings of the Aristotelian society* (pp. 1–21). Oxford, UK: Oxford University Press.

Fiebich, A., & Coltheart, M. (2015). Various ways to understand other minds: Towards a pluralistic approach to the explanation of social understanding. *Mind & Language, 30*, 235–258.

Friedman, O., & Leslie, A. M. (2005). Processing demands in belief–desire reasoning: Inhibition or general difficulty? *Developmental Science, 8*, 218–225.

Gallagher, S. (2015). The problem with 3-year-olds. *Journal of Consciousness Studies, 22*, 160–182.

Gallagher, S., & Hutto, D. D. (2008). Understanding others through primary interaction and narrative practice. In J. Zlatev, T. P. Racine, C. Sinha, & E. Itkonen (Eds.), *Converging evidence in language and communication research (CELCR), Vol. 12: The shared mind: Perspectives on intersubjectivity* (pp. 17–38). Amsterdam: John Benjamins.

German, T. P., & Nichols, S. (2003). Children's counterfactual inferences about long and short causal chains. *Developmental Science, 6*, 514–523.

Goldman, A. I. (2006). *Simulating minds: The philosophy, psychology, and neuroscience of mindreading*. Oxford, UK: Oxford University Press.

Gopnik, A., & Meltzoff, A. N. (1997). *Words, thoughts, and theories*. Cambridge, MA: MIT Press.

Gopnik, A., & Wellman, H. M. (2012). Reconstructing constructivism: Causal models, Bayesian learning mechanisms, and the theory theory. *Psychological Bulletin, 138*, 1085–1108.

Grant, C. M., Riggs, K. J., & Boucher, J. (2004). Counterfactual and mental state reasoning in children with autism. *Journal of Autism and Developmental Disorders, 34*, 177–188.

Guajardo, N. R., Parker, J., & Turley-Ames, K. (2009). Associations among false belief understanding, counterfactual reasoning, and executive function. *British Journal of Developmental Psychology, 27*, 681–702.

Guajardo, N. R., & Turley-Ames, K. J. (2004). Preschoolers' generation of different types of counterfactual statements and theory of mind understanding. *Cognitive Development, 19*, 53–80.

Hansen, M. (2010). If you know something, say something: Young children's problem with false beliefs. *Frontiers in Psychology, 1*. https://doi.org/10.3389/fpsyg.2010.00023.

Harris, P. L., German, T., & Mills, P. (1996). Children's use of counterfactual thinking in causal reasoning. *Cognition, 61*, 233–259.

Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review, 93*, 136–153.

Krzyżanowska, K. (2013). Belief ascription and the Ramsey test. *Synthese, 190*, 21–36.
Kulke, L., & Rakoczy, H. (2018). Implicit theory of mind—An overview of current replications and non-replications. *Data in Brief, 16*, 101–104.
Leahy, B., Rafetseder, E., & Perner, J. (2014). Basic conditional reasoning: How children mimic counterfactual reasoning. *Studia Logica, 102*, 793–810.
Leslie, A. M., & Polizzi, P. (1998). Inhibitory processing in the false belief task: Two conjectures. *Developmental Science, 1*, 247–253.
Lewis, C., & Osborne, A. (1990). Three-year-olds' problems with false belief: Conceptual deficit or linguistic artifact? *Child Development, 61*, 1514–1519.
Lewis, D. (1973). *Counterfactuals*. Oxford, UK: Basil Blackwell.
Low, J., Apperly, I. A., Butterfill, S. A., & Rakoczy, H. (2016). Cognitive architecture of belief reasoning in children and adults: A primer on the two-systems account. *Child Development Perspectives, 10*, 184–189.
Mitchell, P., & Lacohée, H. (1991). Children's early understanding of false belief. *Cognition, 39*, 107–127.
Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods, 7*, 105–125.
Müller, U., Miller, M. R., Michalczyk, K., & Karapinka, A. (2007). False belief understanding: The influence of person, grammatical mood, counterfactual reasoning and working memory. *British Journal of Developmental Psychology, 25*, 615–632.
Musholt, K. (2018). Self and others. *Interdisciplinary Science Reviews, 43*, 136–145.
Nyhout, A., Henke, L., & Ganea, P. A. (2019). Children's counterfactual reasoning about causally overdetermined events. *Child Development, 90*, 610–622.
Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs?. *Science, 308*, 255–258.
Paulus, M., & Sabbagh, M. A. (Eds.). (2018). Understanding theory of mind in infancy and toddlerhood [special issue]. Cognitive Development, 46.
Perner, J. (2005). Infants' insight into the mind: How deep?. *Science, 308*, 214–216.
Perner, J., & Howes, D. (1992). "He thinks he knows": And more developmental evidence against the simulation (role taking) theory. *Mind & Language, 7*, 72–86.
Perner, J., & Roessler, J. (2010). Teleology and causal understanding in children's theory of mind. In J. H. Aguilar & A. A. Buckareff (Eds.), *Causing human actions: New perspectives on the causal theory of action* (pp. 199–228). Cambridge, MA: MIT Press.
Perner, J., Sprung, M., & Steinkogler, B. (2004). Counterfactual conditionals and false belief: A developmental dissociation. *Cognitive Development, 19*, 179–201.
Perner, J., ... Wimmer, H. (1985). "John thinks that Mary thinks that": Attribution of second-order beliefs by 5- to 10-year-old children. *Journal of Experimental Child Psychology, 39*, 437–471.
Peterson, D. M., & Bowler, D. M. (2000). Counterfactual reasoning and false belief understanding in children with autism. *Autism, 4*, 391–405.
Peterson, D. M., & Riggs, K. J. (1999). Adaptive modelling and mindreading. *Mind & Language, 14*, 80–112.
Rafetseder, E., Cristi-Vargas, R., & Perner, J. (2010). Counterfactual reasoning: Developing a sense of "nearest possible world". *Child Development, 81*, 376–389.
Rafetseder, E., Schwitalla, M., & Perner, J. (2013). Counterfactual reasoning: From childhood to adulthood. *Journal of Experimental Child Psychology, 114*, 389–404.
Rasga, C., Quelhas, A. C., & Byrne, R. M. J. (2016). Children's reasoning about other's intentions: False belief and counterfactual conditional inferences. *Cognitive Development, 40*, 46–59.
Riggs, K. J., Peterson, D. M., Robinson, E. J., & Mitchell, P. (1998). Are errors in false belief tasks symptomatic of a broader difficulty with counterfactuality?. *Cognitive Development, 13*, 73–90.
Rubio-Fernández, P., & Geurts, B. (2013). How to pass the false-belief task before your fourth birthday. *Psychological Science, 24*, 27–33.
Ruffman, T. (2014). To belief or not belief: Children's theory of mind. *Developmental Review, 34*, 265–293.
Scanlon, T. M. (1998). *What we owe to each other*. Cambridge, MA: Harvard University Press.
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 of the United States of America, 113*, 13360–13365.
Sodian, B., & Wimmer, H. (1987). Children's understanding of inference as a source of knowledge. *Child Development, 58*, 424–433.
Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science, 18*, 587–592.
Stalnaker, R. C. (1968). A theory of conditionals. In W. L. Harper, R. C. Stalnaker, & G. Pearce (Eds.), IFS. The University of Western Ontario Series in Philosophy of Science (Series of Books in Philosophy of Science, Methodology, Epistemology, Logic, History of Science, and Related Fields, pp. 41–55). Dordrecht, Netherlands: Springer.
Sullivan, K., Zaitchik, D., & Tager-Flusberg, H. (1994). Preschoolers can attribute second-order beliefs. *Developmental Psychology, 30*, 395–402.
Surian, L., Caldi, S., & Sperber, D. (2007). Attribution of beliefs by 13-month-old infants. *Psychological Science, 18*, 580–586.
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. *British Journal of Developmental Psychology, 30*, 172–187.
Van Hoeck, N., Begtas, E., Steen, J., Kestemont, J., Vandekerckhove, M., & Van Overwalle, F. (2014). False belief and counterfactual reasoning in a social environment. *NeuroImage, 90*, 315–325.
Weisberg, D. S., & Gopnik, A. (2013). Pretense, counterfactuals, and Bayesian causal models: Why what is not real really matters. *Cognitive Science, 37*, 1368–1381.
Wellman, H. M. (2014). *Making minds: How theory of mind develops*. Oxford, UK: Oxford University Press.
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.
Westra, E. (2017). Pragmatic development and the false belief task. *Review of Philosophy and Psychology, 8*, 235–257.