http://econtent.hogrefe.com/doi/pdf/10.1027/2151-2604/a000327 - Friday, April 06, 2018 6:19:26 AM - IP Address:139.153.115.11

Original Article

Belief and Counterfactuality

A Teleological Theory of Belief Attribution

Eva Rafetseder[1] and Josef Perner[2,3]

[1]Faculty of Natural Sciences, University of Stirling, UK
[2]Centre for Cognitive Neuroscience, University of Salzburg, Austria
[3]Department of Psychology, University of Salzburg, Austria

Abstract: The development and relation of counterfactual reasoning and false belief understanding were examined in 3- to 7-year-old children (N = 75) and adult controls (N = 14). The key question was whether false belief understanding engages counterfactual reasoning to infer what somebody else falsely believes. Findings revealed a strong correlation between false belief and counterfactual questions even in conditions in which children could commit errors other than the reality bias (r_p = .51). The data suggest that mastery of belief attribution and counterfactual reasoning is not limited to one point in development but rather develops over a longer period. Moreover, the rare occurrence of reality errors calls into question whether young children's errors in the classic false belief task are indeed the result of a failure to inhibit what they know to be actually the case. The data speak in favor of a teleological theory of belief attribution and challenges established theories of belief attribution.

Keywords: counterfactual reasoning, basic conditional reasoning, false belief, teleology-in-perspective

It has been repeatedly reported that young children's answers to counterfactual questions correlate consistently with their answers in the classic false belief test developed by Wimmer and Perner (1983). In the classic false belief task Max places his chocolate into a drawer, but in his absence the chocolate is relocated to the cupboard. Children who understand that Max holds a false belief also predict that he is going to search for his chocolate in the empty drawer. Peterson and Riggs (1999), in their adaptive modeling theory, were the first to propose that understanding false belief depends on counterfactual thinking. Counterfactual thoughts go beyond what we know to be true and allow us to reflect on how else a situation could have unfolded. For example, one may ask where the chocolate would be, if it had not been moved. A number of studies found that answers to counterfactual questions correlate developmentally with answers to the false belief questions, but counterfactual questions tend to be easier. An overview of the correlation coefficients and differences in task difficulty in these studies is shown in Figure 1.

These findings are important as they help us better understand the cognitive basis of belief attribution – as measured by the classic false belief test.[1] As we will argue, the involvement of counterfactual reasoning in belief attribution fits the teleological theory of how belief is understood particularly well. By developing a new test we can evaluate this theory against two existing interpretations of these findings and shed light on basic questions about the nature and development of belief formation. Our data show that the mastery of belief attribution and counterfactual reasoning is not limited to one point in development (i.e., around 4 years) but is more prolonged. Although our work only goes into developments a few years longer (i.e., 7 years) there is evidence that more intricate aspects of counterfactual reasoning might be mastered far into an individual's life.

Rafetseder, Schwitalla, and Perner (2013) illustrated this prolonged development with two conditions: in the 1-puppet condition Carol dirtied the clean floor with her muddy shoes, while in the 2-puppets condition Max and

[1] We assume by default that children become able to represent another person's belief around 4 years when they can provide correct answers to the test question in the classic false belief task. However, use of indirect measures like anticipatory looking (Clements & Perner, 1994; Southgate, Senju, & Csibra, 2007) or longer looking at unexpected outcomes (Onishi & Baillargeon, 2005) indicate a much earlier sensitivity to others' belief around 14–18 months or even earlier. The cognitive basis of these early signs remains hotly disputed (Ruffman, 2014; Setoh, Scott, & Baillargeon, 2016; Wellman, 2014) and has recently been compounded by replication difficulties (Kulke & Rakoczy, 2018). We, therefore, feel confident that passing the classic false belief test indicates an important development, whose relationship to counterfactual thinking is worth investigating.

© 2018 Hogrefe Publishing. Distributed under the
Hogrefe OpenMind License https://doi.org/10.1027/a000001

http://econtent.hogrefe.com/doi/pdf/10.1027/2151-2604/a000327 - Friday, April 06, 2018 6:19:26 AM - IP Address:139.153.115.11

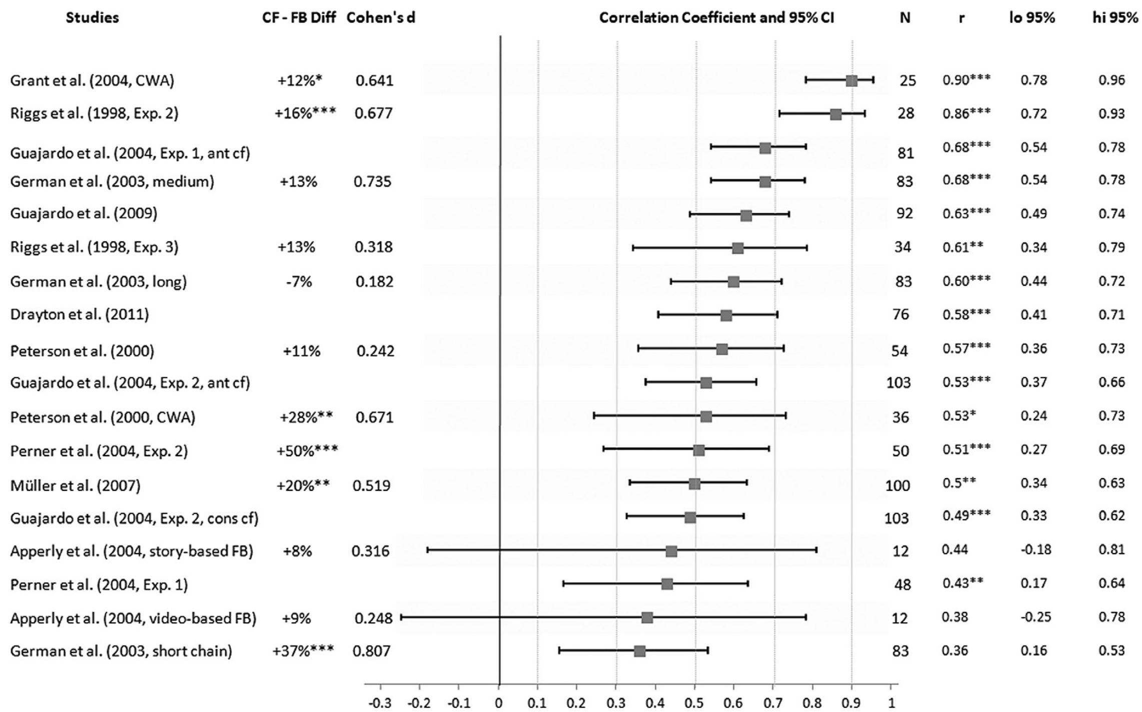

Figure 1. Forest plot for 18 estimates of r – as represented by squares – from 10 papers investigating the correlation of classic false belief tasks with counterfactual tasks. CF-FB Diff shows the mean difference between correct responses to counterfactual (CF) and false belief (FB) questions (sign test: p = .006). 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% CI. Stars represent the significance level of the correlation: *p < .05; **p < .01; ***p < .001.

Carol dirtied the clean floor. In both conditions participants were asked the same counterfactual question, "If Carol had taken her dirty shoes off, would the floor be dirty or clean?" All participants agreed that the floor would be "clean" in the *1-puppet condition*, but in the *2-puppets condition* adults and older children said that the floor would be dirty, while over 80% of the 5- to 6-year-olds concluded the floor to be clean. This finding replicated a previous set of studies (Rafetseder, Cristi-Vargas, & Perner, 2010; Rafetseder & Perner, 2010) and the authors argued that children approach counterfactual questions differently from adults. In whatever way they answer counterfactual questions, their answers should be reflected in their false belief understanding, if these two abilities are linked.

Before expanding on this prediction in more detail, we outline the different theories which we try to assess empirically. Peterson and Riggs (1999) explained the developmental connection between counterfactual reasoning and belief ascription with the use of *modified derivation*, a reasoning strategy considered to support counterfactual and false belief reasoning. To illustrate, consider a boy, Ben, dirtying a floor with his muddy shoes. Avoiding troubles Ben wipes the floor clean and leaves. Unknown to him, his sister Sarah later dirties the clean floor again with her muddy shoes. False belief question: What does Ben think the floor looks like? Counterfactual question: If Sarah had not walked on the floor with her dirty shoes on, what would the floor look like?

Having been told the entire scenario one knows the details of the story (Ben dirtied the floor. He wiped it clean. *Then Sarah dirtied the floor*). These details are stored in one's "database" (Peterson & Riggs' terminology). Since Ben did not see Sarah dirtying the floor, this detail is not stored in Ben's own database (italicized in the example above). When being asked how Ben thinks the floor looks, one must identify differences between one's own and Ben's database and temporarily modify one's own database by ignoring the facts that are not also part of Ben's database. One then poses oneself the question "What does the floor look like?" using the modified database and attribute the resulting answer "clean" to Ben. Similarly, when asked if Sarah had not walked on the floor with her dirty shoes, one is instructed to ignore the fact that *Sarah dirtied the floor* in one's own database. The question "What does the floor look like?" results in "clean," but this time is not attributed to Ben. Not having to attribute the counterfactual answer to another person provided one explanation of why

© 2018 Hogrefe Publishing. Distributed under the
Hogrefe OpenMind License https://doi.org/10.1027/a000001

http://econtent.hogrefe.com/doi/pdf/10.1027/2151-2604/a000327 - Friday, April 06, 2018 6:19:26 AM - IP Address:139.153.115.11

children typically find counterfactual questions easier than false belief questions (see Figure 1, CF-FB Diff). Alternatively, counterfactual questions of this sort may be easier because their antecedent explicitly specifies what one should assume. False belief questions do not offer such specification (Krzýanowska, 2012).

"Inhibition" theory is another theory to explain the developmental correlation between counterfactual tasks and false belief tasks. It assumes a common difficulty of having to inhibit the "lure of the real" (for a discussion, see Robinson & Beck, 2000). In both tasks the correct response involves something that is not the case, for instance, in our footprint example to say that the floor *would be clean* or that Ben *thinks it is clean*, when in reality it has Sarah's footprints on it. It is argued that those 3- to 5-year-old children who do not give the correct answer tend to use "*r*eality *r*easoning" (RR), that is, that (Ben thinks) there are Sarah's footprints on the floor. Robinson and Beck (2000), however, already raised some doubts that difficulties with counterfactual questions are due to a failure to inhibit, because even 3-year-olds are perfectly able to answer conditional questions directed into the future (i.e., future hypotheticals).

Perner and Roessler (2010; Roessler & Perner, 2013) recently provided a new role for counterfactual reasoning in their teleological model of intentional action understanding. We will modify their model into a teleological theory of belief attribution. At the heart of teleology is the conviction that we do not see other agents acting according to some lawful regularities. Instead, we see them as acting for good reasons, which makes their actions rational and meaningful. This works nicely in cases of shared knowledge (agent's database same as own), where no false beliefs are involved. For instance, consider Sarah in the above footprint example. If we learn that she wants to please her mother, who appreciates a clean floor, she has objectively good reasons to clean the floor, because it is dirty. And we may predict that she will likely do so to please her mother. Now consider Ben who also wants to please his mother. He has also objectively good reason to clean the floor, which is dirty. However, we should not predict that he will go and clean it. So his behavior appears to be irrational if he means to please his mother. To rationalize Ben's failure to go clean the floor, Perner and Roessler (2010) suggested that we reason counterfactually: if, as Ben wrongly believes, the floor were clean after he had cleaned it, Ben would have good reasons to not go and clean it again.

Perner and Roessler's use of counterfactual reasoning is designed to preserve rationality of action when agents' view of the world deviates from reality (i.e., one's own view). We want to go one step further and ask whether counterfactual reasoning may also play a role in belief attribution, because we see other agents not only act rationally (for good reasons), but also form rational beliefs.[2] A rational agent should not believe anything but only what he has good reasons to believe. To capture this requirement of rationality we can extend Perner and Roessler's use of counterfactual reasoning for action to belief formation. We do so by using Peterson and Riggs' (1999) notion of an agent's database.

"Database" can be understood as part of an agent's belief system. What we need is a basis of objective events, which give rise to the agent's belief. Perner, Rendl, and Garnham (2007, p. 500) spoke of a "record of tracked events", that is, a record of those events that an agent could track (that happened within his informational, perceptual field). Apperly and Butterfill (2009) coined the term "registration" (of the events that the agent had occasion to register) and Perner and Roessler (2010) of an "experiential record." We, therefore, adopt the following terminological convention: The experiential record for an agent is a record of all events that the agent could register.

All these notions partition the series of all events into those that the agent registered and those he could not. We now can use counterfactual reasoning to determine what, given an agent's specific record of events, he has good reasons to believe: If the actual sequence of events consisted of those that Ben could register (that he wiped his footprints off) ignoring events he did not register (Sarah dirtying the floor), then he would have good reason to not clean the floor again. Now we have a similar, yet different and differently motivated theory from Peterson and Rigg's modified derivation of why belief attribution involves counterfactual reasoning. Inspired by Perner and Roessler we dub it the *teleological theory of belief attribution*.

We construct a test for these theories using the finding by Rafetseder et al. (2013). The development of counterfactual reasoning, as well as belief attribution, has traditionally been investigated with response options limited to the state of reality and the counterfactual or believed state. Under this limitation it seemed that counterfactual reasoning develops between 3 and 5 years as does belief attribution (Wellman, Cross, & Watson, 2001) using reality reasoning (RR) before that age. However, when other plausible response options were made available, much older children had problems providing the correct answers to counterfactual questions in some tasks (Rafetseder et al., 2010, 2013; Rafetseder & Perner, 2010).

[2] Naturally we accept that people have some irrational beliefs at times. But if this happens on a regular basis the person becomes subject to various psychiatric conditions like, suffering from delusions, illusions, paranoia, and so forth.

© 2018 Hogrefe Publishing. Distributed under the Hogrefe OpenMind License https://doi.org/10.1027/a000001

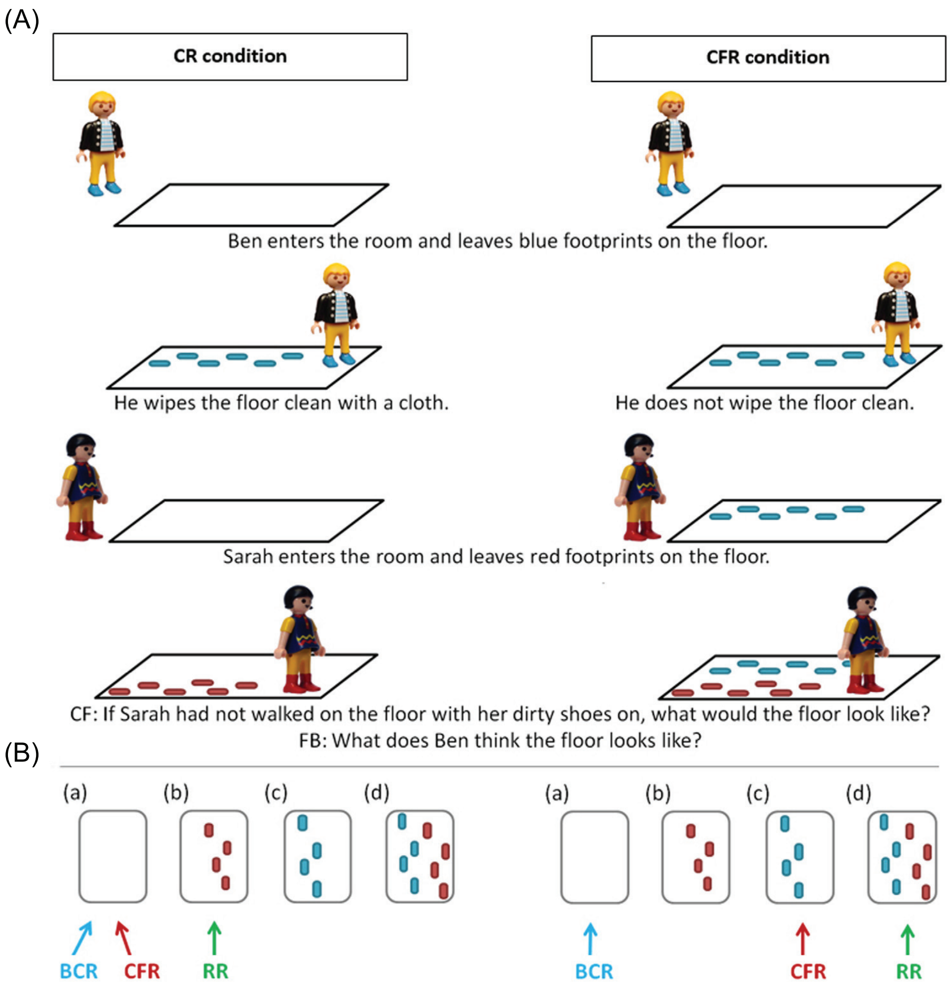

Figure 2. (A) Set up for CR (left) and CFR (right) condition. (B) The resulting answer as determined by different reasoning strategies: RR = reality response; BCR = basic conditional reasoning; CFR = counterfactual reasoning.

Study

The main aim of the current study was to investigate whether the link between false belief and counterfactual questions would still be as close as in previous studies (see Figure 1), if we broadened response options. For example, if errors other than RR were possible, would children respond with the same option to both questions? To elucidate this question we tested 3- to 7-year-old children using a novel false belief task (see Figure 2A), closely aligned to the counterfactual task in Rafetseder et al. (2013) with two conditions. They differ as to which kind of reasoning results in correct answers. Rafetseder et al. (2010) distinguished two types of conditional reasoning: basic conditional reasoning (BCR) and counterfactual reasoning (CFR). In the conditional reasoning (CR) condition both types of reasoning provide the correct answer, while in the counterfactual reasoning (CFR) condition only counterfactual reasoning leads to the correct answer.

For example, a puppet, Ben, enters the room leaving blue footprints on the floor. Ben then either wipes the floor clean (CR condition) or leaves the footprints (CFR condition) before disappearing into the garden. Later Sarah enters the room, leaving red footprints on the floor. Children were asked a counterfactual question (If Sarah had not walked on the floor with her dirty shoes on, what would the floor look like?) and a false belief question (What does Ben think the floor looks like?). Children indicated their answer by pointing at one of the four cards displaying a floor with (a) no footprints, (b) red footprints, (c) blue footprints, or (d) red and blue footprints (Figure 2B). Adults who use counterfactual reasoning (CFR) point at (a) in the CR condition and at (c) in the CFR condition.

Figure 2 indicates the different types of reasoning possible. Of particular interest are RR answers and BCR answers. RR answers occur when children indicate how the floor really looks: option (b) in the CR condition and (d) in the CFR condition. BCR answers occur when, instead

© 2018 Hogrefe Publishing. Distributed under the Hogrefe OpenMind License https://doi.org/10.1027/a000001

http://econtent.hogrefe.com/doi/pdf/10.1027/2151-2604/a000327 - Friday, April 06, 2018 6:19:26 AM - IP Address:139.153.115.11

http://econtent.hogrefe.com/doi/pdf/10.1027/2151-2604/a000327 - Friday, April 06, 2018 6:19:26 AM - IP Address:139.153.115.11

of keeping matters of fact as close as possible to what actually happened (as would CFR), children think they are asked to reason about a new incident with the same characters and environment in which the antecedent holds true (i.e., Sarah took her dirty shoes off). This results in response (a) in both conditions. CFR conditions differ from CR conditions in that only CFR arrives at the correct answer. This difference is important, as the three theories about why false belief attribution relates developmentally to counterfactual reasoning make different predictions. Figure 3 shows hypothetical data patterns predicted by each theory for our paradigm.

According to Peterson and Riggs' (1999) adaptive modeling theory children's answers to counterfactual and false belief questions correlate, because the same modified derivation procedure is needed. Additional factors account for the differences between the tasks. For instance, false belief answers additionally require attribution of the counterfactual answer to somebody else (see Figure 3A). Once children modify their database accordingly (i.e., ignore the fact that Sarah came), they will give the correct answer in the CR and the CFR condition. At present the theory does not provide a plausible account of why modified derivation should be more difficult in one condition than the other.

The teleological theory based on Perner and Roessler (2010; Roessler & Perner, 2013) avoids this problem. It assumes that counterfactual reasoning is an integral part of belief attribution but leaves it open which precise reasoning strategy (BCR or CFR) people use to approach this issue. If CFR is required for a correct answer to the CF question – which is the case in the CFR condition – false belief questions should not be answered correctly before the counterfactual questions. Their correlation thus stays stable (see Figure 3B).

Finally, the inhibition theory (as discussed in Robinson & Beck, 2000) implies that counterfactual and false belief questions should correlate for the younger children who commit the RR error. This theory does not specify why the false belief question tends to be harder than the counterfactual question in existing data and it does not include predictions for BCR errors. So, no prediction follows for relative difficulty (see Figure 3C).

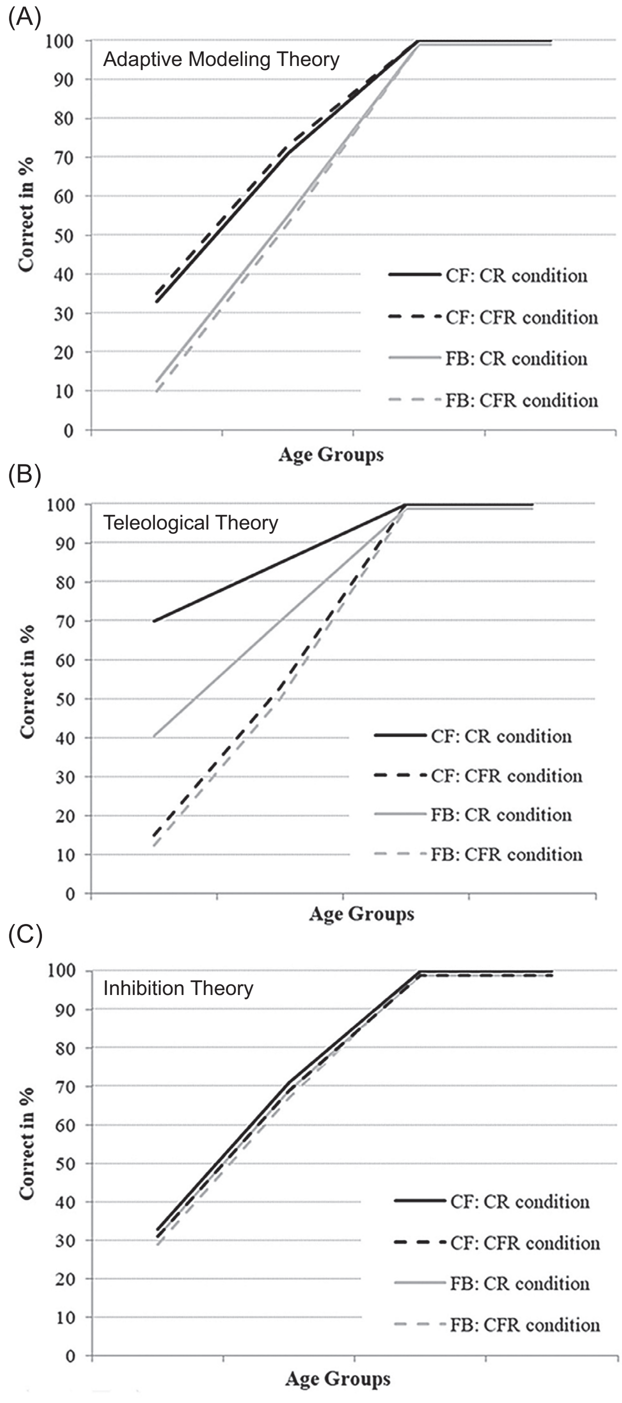

Figure 3. Hypothetical data pattern according to (A) Adaptive Modeling Theory, (B) Teleological Theory, and (C) Inhibition Theory.

Method

Participants

A total of 75 children between 38 and 92 months and 13 adults between 19 and 57 years (M_{age} = 32.54 years, SD = 14.60, 7 females) participated in this study. Adults were included to specify the answers that mature counterfactual reasoners would give in our paradigm. Testing occurred in a separate room and some children expressed a desire to leave the room before data collection could be finished (n = 4). Six further children were tested, but excluded, for incorrectly answering one or more C2 control questions. The final sample consisted of 65 children

© 2018 Hogrefe Publishing. Distributed under the Hogrefe OpenMind License https://doi.org/10.1027/a000001

http://econtent.hogrefe.com/doi/pdf/10.1027/2151-2604/a000327 - Friday, April 06, 2018 6:19:26 AM - IP Address:139.153.115.11

(M_{age} = 62.43 months, SD = 14.44, 35 girls) who were predominantly from a middle-class background and were recruited from playgroups in the central belt of Scotland (n = 48) and in Germany (n = 17). Parents gave written consent for their children to participate and the children each gave their assent to take part in the tasks. Ethical approval was granted for this research by the General University Ethics Panel of the University of Stirling (Approval Reference: GUEP 150).

Design

All participants completed four stories, each of whom asked a false belief and a counterfactual question, alongside a classic false belief task (Wimmer & Perner, 1983). The order of the four stories followed a Latin Square Design and the classic false belief task was always presented midway. The order of the test questions and the order of the two conditions CR and CFR were fully counterbalanced within participants. To prevent any bias resulting from right- or left-handedness, the four answer cards were presented randomly.

Materials and Procedure

Participants received a total of four stories using Playmobil© puppets to visually support the narration.

Footprint Scenario

Participants were introduced to a boy, Ben, who had been playing with blue crayon. Ben entered the room (represented with a white square) and left blue footprints on the floor. He either wiped the floor clean with a cloth (CR) or left the blue footprints unclean (CFR), before heading back outside. The puppet left the scene entirely and participants were asked C1 “Are there footprints on the floor or no footprints on the floor?” All participants correctly answered, depending on the condition, with “no footprints” (CR) or with “footprints” (CFR). Then a girl, Sarah, entered the scene, who had been playing with red crayon, and she left red footprints on the floor. To control for any false assumptions participants were asked C2 “Did Ben see Sarah leaving red footprints on the floor?” Those participants who answered “yes” were excluded from the sample. Finally, with the red footprints (and the blue footprints in CFR) still visible, participants were asked two test questions in a counterbalanced order, “What does Ben think the floor looks like? Does he think there are footprints or no footprints?” (False belief question) and “If Sarah had not walked on the floor with her dirty shoes on, what would the floor look like? Would there be footprints on the floor or would there be no footprints on the floor?” (Counterfactual question). In between the two questions participants were asked a reality question “How does the floor actually look?” To answer these questions, participants were asked to point at one of four randomly laid out cards, displaying a floor with (a) no footprints, (b) red footprints, (c) blue footprints, or (d) red and blue footprints.

Waterpark Scenario

Anna was seen playing with her blue ball in the grass. She either tidied it away (CR) or left it in the grass (CFR) before going for a swim in the pool. Then Mark entered the scene and played with a green ball in the grass.

Whiteboard Scenario

Tom was seen drawing a house on a whiteboard with a green pen. He either wiped it off (CR) or left it (CFR) before he went to play in a different room. Then Sally entered the scene and drew a red flower on the whiteboard.

Tablecloth Scenario

The scene showed a table with a white surface when a girl, Claire, entered and placed her red lemonade on the table, leaving a big red stain. She either wiped the table clean (CR) or she left the red stain (CFR) before heading out to visit a friend. Then Jack entered and put his yellow lemonade on the table causing a yellow stain on the surface.

Classic False Belief Task

A classic false belief task (Wimmer & Perner, 1983) was presented on a laptop using PowerPoint© (Perner, Mauer, & Hildenbrand, 2011). It showed Lisa playing with her teddy before placing it in a red basket and leaving for the kitchen. While away, her brother Tom moved her teddy from the red basket into a yellow basket, before heading outside to play. Participants were then asked five control questions: (1) Where did Lisa put her teddy? (2) Where is teddy now? (3) Who put teddy there? (4) Did Lisa see that? (5) Where did Lisa put her teddy at the beginning? If any of these questions were answered incorrectly (n = 2), the PowerPoint was restarted. Finally, Lisa reentered the scene, and participants were asked, “Where will Lisa search *first* for her teddy?” Lisa was then shown to search the red basket and participants were asked, “Why did Lisa search for her teddy in the red basket?”

Scoring

Each participant answered a total of four counterfactual and four false belief questions, two of each in CR and two in CFR. Participants scored one point if they chose answer card (a) in the CR condition and answer card (c) in the CFR condition. This resulted in a maximum score of 2 for each type of question in each condition.

In the classic false belief task, participants gained one point for answering the prediction question with “red.” The explanation question was scored based on Perner, Lang, and Kloo (2002; see also Priewasser,

© 2018 Hogrefe Publishing. Distributed under the Hogrefe OpenMind License https://doi.org/10.1027/a000001

http://econtent.hogrefe.com/doi/pdf/10.1027/2151-2604/a000327 - Friday, April 06, 2018 6:19:26 AM - IP Address:139.153.115.11

Roessler, & Perner, 2013). Participants scored two points if they ascribed a mental state to Lisa, for instance, "She *thinks* it is still there," or "She *didn't know* it was moved." One point was obtained for identifying important facts about the story, for example, "That's where she put it before it was moved," or "She put it there first." No points were awarded when the participants used irrelevant facts about the story or did not give a verbal response. Finally, a sum false belief score was calculated (Max_{corr} = 3).

Results

Adult Controls

Adult controls showed near ceiling performance, with no errors in the CR condition, and two errors (by selecting the a-card) in the CFR condition, one for each question type, false belief, and counterfactual question. With regard to the classic false belief task, all adult controls answered the prediction question correctly. They received a sum score of 2.343 (SD = 0.54). Because of the near ceiling performance, no further analysis was conducted.

Preliminary Analyses

The story material did not affect children's answers to counterfactual questions, Cochran's $\chi^2(65) = 5.21$, $p = .16$, or their answers to false belief questions, Cochran's $\chi^2(65) = 4.44$, $p = .22$. There were no effects of gender, neither on the answers to the two question types in the two conditions (all ps > .45) nor on the mean classic false belief score ($p = .71$). Therefore, these variables were not considered further.

Main Analyses

Mean performance on test questions was examined using a mixed factorial repeated measures analysis of variance (ANOVA) with question type (false belief vs. counterfactual) and required reasoning (CR vs. CFR) as within subject factors and age (4 levels: 3-year-olds: M_{age} = 44.31; 4-year-olds: M_{age} = 55.81; 5-year-olds: M_{age} = 67.69; 6-year-olds: M_{age} = 80.76) as a between subject factor. A main effect of question type, $F(1, 61) = 6.83$, $p = .011$, $\eta^2 = .10$, a main effect of required reasoning, $F(1, 61) = 27.59$, $p < .001$, $\eta^2 = .31$, and a main effect of age, $F(1, 61) = 10.35$, $p < .001$, $\eta^2 = .34$, were obtained. The interaction between question type and required reasoning was marginally significant, $F(1, 61) = 3.15$, $p = .081$, $\eta^2 = .05$. While counterfactual (M_{corr} = 1.10, SE = .096) and false belief questions (M_{corr} = 1.04, SE = .093) were answered at a comparable level in the CFR condition, $F(1, 61) = .45$, $p = .50$, $\eta^2 = .007$, children answered more counterfactual questions (M_{corr} = 1.63, SE = .077) accurately than false belief questions (M_{corr} = 1.33, SE = .09) in the CR condition,

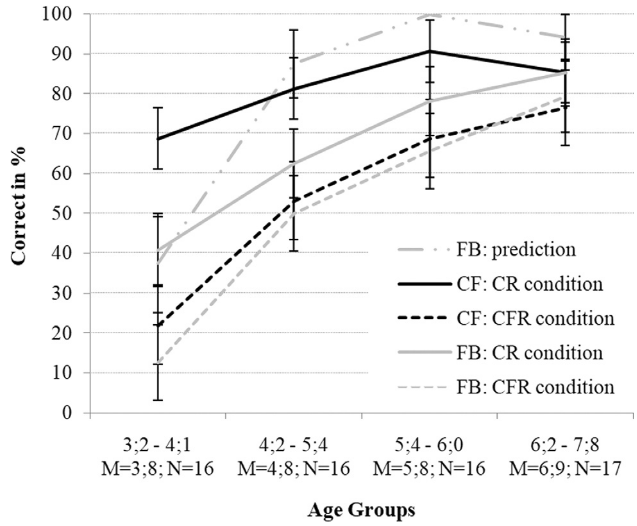

Figure 4. Mean number of correct responses to counterfactual and false belief questions in each condition (CR and CFR) as well as the prediction question of the classic false belief tasks displayed separately for age.

$F(1, 61) = 9.52$, $p = .003$, $\eta^2 = .14$ (see Figure 4). The data pattern found corresponds particularly well with the hypothetical data pattern (B) in Figure 2.

Required reasoning also interacted marginally significantly with age, $F(1, 61) = 2.58$, $p = .061$, $\eta^2 = .11$. We found an age effect in both conditions CR, $F(1, 61) = 4.75$, $p = .005$, $\eta^2 = .19$, and CFR, $F(1, 61) = 10.37$, $p < .001$, $\eta^2 = .34$. However, as Table 1 shows, all questions correlated with age except for counterfactual questions in the CR condition, evidently because even the youngest children had little problems with this question. When age was controlled, correlations between counterfactual and false belief questions were still moderate to large (bracketed numbers).

Analysis of Reasoning Strategies

Which reasoning strategy children used can be identified by their response pattern for the CR and the CFR condition as shown in Figure 5. This can be done for the counterfactual question and the false belief question for the first and the second pair of tasks.

Figure 6 displays the contingency of reasoning strategies for answering the counterfactual and the belief question separated for the first (Figure 6A) and the second pair of stories (Figure 6B). The contingency between strategies used to answer the counterfactual question and the belief question was high for the first ($\Phi = .66$, $p = .001$) as well as the second pair of stories ($\Phi = .83$, $p < .001$). Also the contingency of strategies used in the first pair and the second pair of stories was high for the counterfactual questions, $\Phi = .72$, $p < .001$, as it was for false belief questions, $\Phi = .69$, $p < .001$.

© 2018 Hogrefe Publishing. Distributed under the Hogrefe OpenMind License https://doi.org/10.1027/a000001

http://econtent.hogrefe.com/doi/pdf/10.1027/2151-2604/a000327 - Friday, April 06, 2018 6:19:26 AM - IP Address:139.153.115.11

Table 1. Correlations [partial correlations controlling for age] between age, counterfactual questions in the CR condition (CF_{CR}) and in the CFR condition (CF_{CFR}), false belief questions in the CR condition (FB_{CR}) and in the CFR condition (FB_{CFR}), and the classic false belief question (prediction and sum score)

	Counterfactual question		False belief question		False belief classic	
Variable	CR	CFR	CR	CFR	Prediction	Sum
Age	.19	.48***	.48***	.56***	.50***	.46***
CF_{CR}		.42***	.37**	.31**	.38**	.39***
CF_{CFR}	[.38]**		.52***	.64***	.38**	.30*
FB_{CR}	[.32]*	[.37]**		.49***	.59***	.53***
FB_{CFR}	[.25]*	[.51]***	[.30]*		.37**	.39**
$FB_{classic}$ prediction	[.34]**	[.18]	[.45]***	[.13]		.77***
$FB_{classic}$ sum	[.34]**	[.10]	[.40]***	[.18]	[.71]***	

Notes. CF = counterfactual question, CR = conditional reasoning, CFR = counterfactual reasoning, FB = False belief question. *$p < .05$; **$p < .01$; ***$p \leq .001$ (two-tailed).

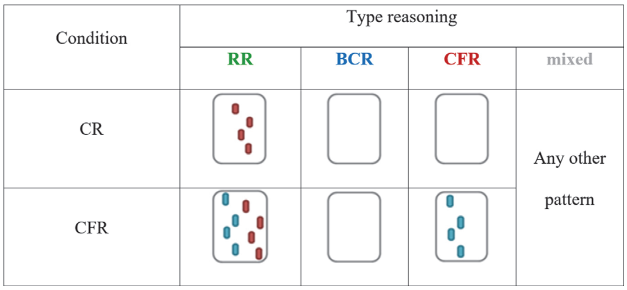

Figure 5. Pattern of responses over the two conditions indicative of a particular reasoning strategy.

Figure 6 shows particularly clearly that children capable of CFR used it for both counterfactual and false belief questions (first: n = 19; second: n = 28). Children who applied BCR on the false belief question either used BCR (first: n = 6; second: n = 6) or CFR (first: n = 5; second: n = 6) on the counterfactual question. Other than that, use of one strategy for one question could go with any other strategy for the other question.

Relation Between Classic and Footprint False Belief Task

The prediction question of the classic false belief question correlated significantly higher with the false belief question in the CR condition (r = .45, p < .001) than in the CFR condition (r = .13, p > .05), z = 2.35, p = .019.[3] Yet, the classic false belief question was significantly easier than the false belief question in the CR condition, $t(64)$ = 2.96, p = .004, presumably due to the higher number of response options (4) in the false belief task of the CR condition than the classic task (2).

Discussion

Results from the CR condition confirm existing findings (Figure 1): children gave more correct answers to the counterfactual than the false belief question and their answers to these two questions were correlated. The comparison between the CR and CFR condition also replicated previous findings (Rafetseder et al., 2013) of CR conditions being easier than CFR conditions. However, this difference is not as large as in the original studies because children found our version of the footprint CFR condition easier (50% correct around 5 years) than the one used by Rafetseder et al. (2013, Figure 2: 50% correct around 9 years). Two changes can plausibly explain that difference: the differently colored footprints may have made it easier to realise that if Sarah had taken off her shoes then only the red footprints would be missing and not also Ben's blue ones. Whereas in the original study the two characters made the same color footprints which may have trapped children more easily into the typical BCR strategy of reasoning, "if Sarah takes off her dirty shoes then no dirty footprints would be added." Indeed the BCR errors were much more pronounced in that study (on average 82% of the 5- to 6-year-olds) than in the present study (around 19% in the same age range). Another reason for the better performance in the CFR condition is the fact that Ben dirtied the floor before Sarah did and before Sarah had entered the scene. Hence, the counterfactual antecedent about Sarah having taken off her shoes could not have had any causal influence on Ben's earlier behavior. In the original study Sarah dirtied the floor first. This made it possible to assume that if she had taken off her shoes then Ben would have been faced with a clean floor and might also have

[3] The z-score is the result of a test of the equality of two correlation coefficients obtained from the same sample, with the two correlations sharing one variable ($FB_{classic}$), as described by Lee and Preacher (2013).

© 2018 Hogrefe Publishing. Distributed under the Hogrefe OpenMind License https://doi.org/10.1027/a000001

http://econtent.hogrefe.com/doi/pdf/10.1027/2151-2604/a000327 - Friday, April 06, 2018 6:19:26 AM - IP Address:139.153.115.11

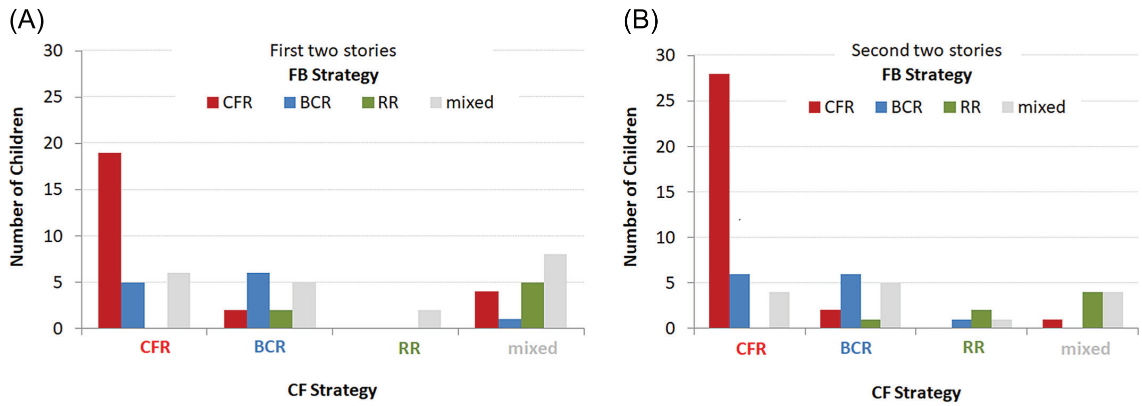

Figure 6. Contingency between strategy used for the CF question (*x*-axis) and for the FB question (color of bar). Strategies were determined separately for the first two stories (one CR and one CFR story) and the second two stories (CR and CFR).

taken off his shoes. With that assumption even CFR reasoners would answer with "clean," which was classified as a wrong response indicating the use of BCR. In fact, Nyhout, Henke, and Ganea (in press) showed that this reversal of who dirtied the floor first had exactly this effect.

The central and important new finding is the strong correlation between answers to the counterfactual and false belief questions in the CFR condition. The correlation was at least as strong (r_p = .51) as in the CR condition (r_p = .32). This supports the idea that children's counterfactual reasoning strategy is an integral part of belief attribution as suggested by the teleological theory. The modified derivation theory has difficulty accounting for this because modified derivation gives the same result for CFR as well as CR problems. Unless the theory can provide a plausible explanation of why modified derivation should be more difficult in CFR than in CR the theory cannot account for our data. Also the inhibition theory has difficulty accounting for this prolonged relationship between belief attribution and counterfactual reasoning. It drew its explanatory strength from the fact that the young children in the original studies found reality-based answers difficult to inhibit leading to many RR errors. In the present study RR errors play practically no role for counterfactual questions (Figure 6) and also only a limited role for false belief questions. Hence problems inhibiting reality-based responses can neither account for the difficulty of the CFR condition nor the strong correlation between answers to counterfactual and false belief questions in this condition. To explain the data the theory would need to claim that BCR responses are difficult to inhibit. But why should that be, and why should they be even more difficult to inhibit than reality-based responses? With reference to Figure 4, inhibition cannot explain why the FB question in the CR condition is more difficult than the CF question, and why both questions are more difficult in the CFR than in the CR condition. So, we see that from the three theories that have addressed the relationship between belief attribution and counterfactual reasoning teleology can most easily accommodate our data.

Apart from this specific contribution our data also speak to wider issues of belief ascription. A persistent developmental question is whether children acquire a competent understanding of belief at a particular point in development or whether this acquisition is a prolonged and gradual process. For instance, we know that children's performance on false belief tasks is influenced by factors like the strength of the hindsight bias and their ability to inhibit such a bias (e.g., Bernstein, Atance, Meltzoff, & Loftus, 2007). Our data suggest in addition that even the determination of the content of beliefs develops gradually depending on the kind of counterfactual reasoning needed. This question whether the concept of belief is acquired as a package (Rakoczy, Bergfeld, Schwarz, & Fizke, 2015) or incrementally (Perner, Huemer, & Leahy, 2015) is also an issue for second-order belief attribution and the understanding of the "intensionality"[4] of belief.

Another wider issue concerns the question whether our data provide evidence for teleology-in-perspective as the

[4] Intensionality is the technical term for the fact that co-referential terms cannot be replaced in sentences about belief without potentially affecting the truth of the sentence, for example, Heinz sees a die, which happens to be made of rubber and serves as an eraser. By just looking at it, Heinz cannot tell that it is also an eraser. Hence, one cannot rephrase "Heinz knows where the die is," as, "Heinz knows where the eraser is," even though "the die" and "the eraser" do refer to the same object.

© 2018 Hogrefe Publishing. Distributed under the Hogrefe OpenMind License https://doi.org/10.1027/a000001

http://econtent.hogrefe.com/doi/pdf/10.1027/2151-2604/a000327 - Friday, April 06, 2018 6:19:26 AM - IP Address:139.153.115.11

way we understand other agents' beliefs in contrast to the entrenched duo of theory use and simulation. Teleology's hallmark is to reintroduce the insight from action theory (Anscombe, 1957) that we understand agents as acting for good reasons (Perner & Roessler, 2010) and not just according to some lawful causal regularities as featured in theory theory (e.g., Gopnik & Meltzoff, 1998). Teleology's weakness is to account for belief-based actions, where agents act on reasons within their subjective perspective. To avoid the technique of imaginative identification propounded by simulation theory, teleology provides counterfactual reasoning about reasons for action and for belief (novel contribution) for this purpose. The question then is whether the present data also favor teleology over theory use and simulation.

To see why theory theory has no natural place for counterfactual reasoning the following typical theory theory account of the classic false belief scenario will help (Carruthers, 2013, p. 160): "Consider a false-belief task of the sort presented to infants: 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. ... noting 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. ... now judging the content, *the agent thinks: the doll is in the green box*." Again it is difficult to see what use could be made here of counterfactual reasoning, which raises the question why belief attribution should emerge in tune with counterfactual reasoning.

Simulation might provide a role for counterfactual reasoning in two places, shifting from one's own perspective to the simulated agent's perspective or when reasoning within the other's perspective. We can rule out the latter because Ben does not engage in counterfactual reasoning and, therefore, there is no use for counterfactual reasoning once one has assumed Ben's perspective. Would counterfactual reasoning be required to shift to Ben's perspective? It seems superfluous. If one knows that Ben has not seen Sarah dirty the floor then one imagines the world from Ben's perspective as not including Sarah's footprints. Again, it is not clear what role counterfactual reasoning could play in this process either.

Simulation theory might, however, have a use for counterfactual reasoning on some occasions to explain recalcitrant data. Indeed Short and Riggs (2016) recently defended simulation theory against Saxe's (2005) claim that it could not explain why children assume that not knowing leads to getting it wrong (Ruffman, 1996). In their defense of simulation Short and Riggs used adaptive modeling in combination with specific assumptions of how people represent disjunctive knowledge (the bead is either red or green). Nevertheless, this elegant move does not show that counterfactual reasoning is a natural partner for simulation theory.

As we have just seen in the case of simulation theory, counterfactual reasoning can be given a role to account for particular data. Yet, neither theory theory nor simulation theory motivates the need for counterfactual reasoning in general. By contrast, teleology needs to preserve the view that mistaken agents have rational beliefs and act for good reasons. For this, teleology depends on counterfactual reasoning.

Acknowledgments

We are grateful to the children and their parents who participated in this study. We thank managers of Castlehill Nursery, Glasgow University Nursery, and University Stirling Kindergarten as well as Raenbow Theatre Production Glasgow and the Kinderhaus Konstanz, Germany. Special thanks are owed to Katherine Reid and Christine McMahon for their unremitting effort to collect the data.

This research was funded by the Austrian Science Fund "Rule-understanding, subjective perspectives, and social display rules" (FWF I 637-G15), by the German Science Foundation through a Mercator-Fellowship with the grant SP 279/18-2, and by a Summer Bursary of the University of Stirling's Psychology Division.

References

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

Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review, 116*, 953–970. https://doi.org/10.1037/a0016923

Apperly, I. A., Samson, D., Chiavarino, C., & Humphreys, G. W. (2004). Frontal and temporo-parietal lobe contributions to theory of mind: Neuropsychological evidence from a false-belief task with reduced language and executive demands. *Journal of Cognitive Neuroscience, 16*, 1773–1784. https://doi.org/10.1162/0898929042947928

Bernstein, D. M., Atance, C., Meltzoff, A. N., & Loftus, G. R. (2007). Hindsight bias and developing theories of mind. *Child Development, 78*, 1374–1394. https://doi.org/10.1111/j.1467-8624.2007.01071.x

Carruthers, P. (2013). Mindreading in infancy. *Mind & Language, 28*, 141–172. https://doi.org/10.1111/mila.12014

Clements, W. A., & Perner, J. (1994). Implicit understanding of belief. *Cognitive Development, 9*, 377–397. https://doi.org/10.1016/0885-2014(94)90012-4

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. https://doi.org/10.1016/j.jecp.2010.09.007

German, T. P., & Nichols, S. (2003). Children's counterfactual inferences about long and short causal chains. *Developmental Science, 6*, 514–523. https://doi.org/10.1111/1467-7687.00309

© 2018 Hogrefe Publishing. Distributed under the Hogrefe OpenMind License https://doi.org/10.1027/a000001

http://econtent.hogrefe.com/doi/pdf/10.1027/2151-2604/a000327 - Friday, April 06, 2018 6:19:26 AM - IP Address:139.153.115.11

Gopnik, A., & Meltzoff, A. (1998). *Words, thoughts, and theories (Learning, development, and conceptual change)*. Cambridge, MA: MIT Press.

Grant, C. M., Riggs, K. J., & Boucher, J. (2004). Counterfactual and mental state reasoning in children with autism. *Journal of Autism and Developmental Disorder, 34*, 177–188. https://doi.org/10.1023/B:JADD.0000022608.57470.29

Guajardo, N. R., Parker, J., & Turley-Ames, K. J. (2009). Associations among false belief understanding, counterfactual reasoning, and executive function. *British Journal of Developmental Psychology, 27*, 681–702. https://doi.org/10.1348/026151008X357886

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. https://doi.org/10.1016/j.cogdev.2003.09.002

Kulke, L., & Rakoczy, H. (2018). Implicit Theory of Mind – An overview of current replications and non-replications. *Data in brief, 16*, 101–104. https://doi.org/10.1016/j.dib.2017.11.016

Krzýanowska, K. (2012). Belief ascription and the Ramsey Test. *Synthese, 190*, 21–36. https://doi.org/10.1007/s11229-012-0160-5

Lee, I. A., & Preacher, K. J. (2013, September). *Calculation for the test of the difference between two dependent correlations with one variable in common*. [Computer software]. Retrieved from http://quantpsy.org

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. https://doi.org/10.1037/1082-989X.7.1.105

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. https://doi.org/10.1348/026151007X182962

Nyhout, A., Henke, L., & Ganea, P. A. (in press). Children's counterfactual reasoning about causally overdetermined events. *Child Development*. https://doi.org/10.1111/cdev.12913

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science, 308*, 255–258. https://doi.org/10.1126/science.1107621

Perner, J., Huemer, M., & Leahy, B. (2015). Mental files and belief: A cognitive theory of how children represent belief and its intensionality. *Cognition, 145*, 77–88. https://doi.org/10.1016/j.cognition.2015.08.006

Perner, J., Lang, B., & Kloo, D. (2002). Theory of mind and self-control: More than a common problem of inhibition. *Child Development, 73*, 752–767. https://doi.org/10.1111/1467-8624.00436

Perner, J., Mauer, M. C., & Hildenbrand, M. (2011). Identity: Key to children's understanding of belief. *Science, 333*, 474–477. https://doi.org/10.1126/science.1201216

Perner, J., Rendl, B., & Garnham, A. (2007). Objects of desire, thought, and reality: Problems of anchoring discourse referents in development. *Mind & Language, 22*, 475–513. https://doi.org/10.1111/j.1468-0017.2007.00317.x

Perner, J., & Roessler, J. (2010). Teleology and causal reasoning in children's Theory of Mind. In J. H. Aguilar & A. A. Buckareff (Eds.), *Causing human action: New perspectives on the causal theory of action* (pp. 199–228). Cambridge, MA: Bradford Book/MIT Press.

Perner, J., Sprung, M., & Steinkogler, B. (2004). Counterfactual conditionals and false belief: A developmental dissociation. *Cognitive Development, 19*, 179–201. https://doi.org/10.1016/j.cogdev.2003.12.001

Peterson, D. M., & Bowler, D. M. (2000). Counterfactual reasoning and false belief understanding in children with autism. *Autism, 4*, 391–405. https://doi.org/10.1177/1362361300004004005

Peterson, D. M., & Riggs, K. J. (1999). Adaptive modelling and mindreading. *Mind & Language, 14*, 80–112. https://doi.org/10.1111/1468-0017.00104

Priewasser, B., Roessler, J., & Perner, J. (2013). Competition as rational action: Why young children cannot appreciate competitive games. *Journal of Experimental Child Psychology, 116*, 545–559. https://doi.org/10.1016/j.jecp.2012.10.008

Rafetseder, E., Cristi-Vargas, R., & Perner, J. (2010). Counterfactual reasoning: Developing a sense of "nearest possible world". *Child Development, 81*, 376–389. https://doi.org/10.1111/j.1467-8624.2009.01401.x

Rafetseder, E., & Perner, J. (2010). Is reasoning from counterfactual antecedents evidence for counterfactual reasoning? *Thinking and Reasoning, 16*, 131–155. https://doi.org/10.1080/13546783.2010.488074

Rafetseder, E., Schwitalla, M., & Perner, J. (2013). Counterfactual reasoning: From childhood to adulthood. *Journal of Experimental Child Psychology, 114*, 389–404. https://doi.org/10.1016/j.jecp.2012.10.010

Rakoczy, H., Bergfeld, D., Schwarz, I., & Fizke, E. (2015). Explicit theory of mind is even more unified than previously assumed: Belief ascription and understanding aspectuality emerge together in development. *Child Development, 86*, 486–502. https://doi.org/10.1111/cdev.12311

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. https://doi.org/10.1016/S0885-2014(98)90021-1

Robinson, E. J., & Beck, S. (2000). What is difficult about counterfactual reasoning? In P. Mitchell & K. J. Riggs (Eds.), *Children's reasoning and the mind* (pp. 101–119). Hove, UK: Psychology Press.

Roessler, J., & Perner, J. (2013). Teleology: Belief as perspective. In S. Baron-Cohen, M. Lombardo, & H. Tager-Flusberg (Eds.), *Understanding other minds: Perspectives from developmental social neuroscience* (pp. 35–51). Oxford, UK: Oxford University Press.

Ruffman, T. (1996). Do children understand the mind by means of a simulation or a theory? Evidence from their understanding of inference. *Mind & Language, 11*, 388–414. https://doi.org/10.1111/j.1468-0017.1996.tb00053.x

Ruffman, T. (2014). To belief or not belief: Children's theory of mind. *Developmental Review, 34*, 265–293. https://doi.org/10.1016/j.dr.2014.04.001

Saxe, R. (2005). Against simulation: The argument from error. *Trends in Cognitive Sciences, 9*, 174–179. https://doi.org/10.1016/j.tics.2005.01.012

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*, 13360–13365. https://doi.org/10.1073/pnas.1609203113

Short, T. L., & Riggs, K. J. (2016). Defending simulation theory against the argument from error. *Mind & Language, 31*, 248–262. https://doi.org/10.1111/mila.12103

Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science, 18*, 586–592. https://doi.org/10.1111/j.1467-9280.2007.01944.x

Wellman, H. M. (2014). *Making minds: How theory of mind develops*. Ann Arbor, MI: University of Michigan.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: the truth about false belief. *Child*

© 2018 Hogrefe Publishing. Distributed under the Hogrefe OpenMind License https://doi.org/10.1027/a000001

Development, 72, 655–684. https://doi.org/10.1111/1467-8624.00304
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. https://doi.org/10.1016/0010-0277(83)90004-5

Received August 18, 2017
Revision received December 23, 2017
Accepted December 23, 2017
Published online March 14, 2018

Eva Rafetseder
Faculty of Natural Sciences
University of Stirling
Stirling FK9 4LA
UK
eva.rafetseder@stir.ac.uk

http://econtent.hogrefe.com/doi/pdf/10.1027/2151-2604/a000327 - Friday, April 06, 2018 6:19:26 AM - IP Address:139.153.115.11

© 2018 Hogrefe Publishing. Distributed under the Hogrefe OpenMind License https://doi.org/10.1027/a000001