

## Counterfactual Reasoning and Reasoning with Beliefs

Similar steps or separate paths?

Eva Rafetseder\* <sup>1</sup>

Christine Hofer<sup>1</sup>

Sabrina Ecker<sup>1</sup>

Josef Perner<sup>1, 2</sup>

<sup>1</sup>Department of Psychology, University of Salzburg, Austria

<sup>2</sup>Centre for Neurocognitive Research, University of Salzburg, Austria

\* Corresponding author.

Department of Psychology, University of Salzburg

Hellbrunnerstraße 34, A-5020 Salzburg, Austria

Tel.: +43-662-8044-5109

Fax: +43-662-8044-5126

E-mail address: [eva.rafetseder@sbg.ac.at](mailto:eva.rafetseder@sbg.ac.at)

### Abstract

Riggs, Peterson, Robinson, and Mitchell (1998) claimed that children have to apply counterfactual reasoning (CFR) when they answer false-belief questions. However, recent findings indicate that children's answers to counterfactual questions may not be based on CFR but they give the correct answers using basic conditional reasoning (BCR: Rafetseder, Cristi-Vargas, & Perner, 2010), i.e., applying universally quantified conditionals that express general regularities. Use of CFR tends to emerge not until 9 to 12 years.

Two studies investigated how children perform on false-belief tasks when controlling for answers based on BCR. Both studies found a highly significant correlation ( $r = .58$ ) between children using CFR to answer the counterfactual question and their correct answers to the false belief question. Moreover, hardly any children gave correct answers to the belief and wrong answers to the counterfactual question, suggesting that counterfactual reasoning is a prerequisite for predicting actions based on false beliefs.

*Keywords.* Counterfactual reasoning; Basic conditional reasoning; False Belief;

Children take a very important step towards social understanding when they are around 4 years old. They display an understanding that other people's beliefs can be different from their own. This has been put to test with the "false-belief" task. In the original version an unexpected change happens in the absence of a person: Maxi puts chocolate into the table drawer (location 1). He then goes to the playground. In his absence, his mother uses a piece of chocolate for her cake but then does not put it back into the drawer but puts it into the cupboard (location 2). Children are then asked where Maxi will search for his chocolate when he comes back. Only children who understand that Maxi holds a false belief will point correctly to the drawer (Wimmer & Perner, 1983). A meta-analysis (Wellman, Cross, & Watson, 2001; 178 studies) revealed that performance on several variants of this task lies below chance when children are 41 months old but reaches above chance when children are 48 months old. The authors ruled out that this important developmental change is affected by the surface features of the tasks, such as type of tasks (unexpected location vs. unexpected content) or type of question (e.g., "Where will Maxi look?" vs. "What does Maxi think?").

Perner (2000) gave a concise overview of different positions about what is accounting for this developmental change. One of these followed the argument by Peterson & Riggs (1999) that counterfactual reasoning is a necessary ingredient to passing a false-belief task. This claim is based on a study by Riggs, Peterson, Robinson, and Mitchell (1998), who found that performances on false belief tasks and on counterfactual tasks are highly correlated ( $r = .86, p < .001$ , study 2). They used stories such as the Maxi-story described above and asked a false-belief question ("Where will Maxi look for his chocolate?") and a counterfactual question ("If mother had not baked a cake, where would the chocolate be?"). Although performance was highly correlated, children solved significantly more counterfactual questions (59%) than false belief questions (43%) which is compatible with the suggestion that counterfactual reasoning is a prerequisite for understanding false belief based actions.

Peterson and Riggs (1999) deduced from the logical structure of the tasks that one of the enabling conditions to succeed is the ability to apply modified derivation (MD). Fully competent reasoners as well as children who fail on both tasks are aware of the fact that at time  $t_1$  the chocolate was put into location 1 (by Maxi) and at time  $t_2$  it was put into location 2 (by Maxi's mum). Both groups are also aware of the norm that things usually stay where they are. What makes up for the difference in performance seems to be that fully competent reasoners modify their answer-derivation process: they ignore the fact that mother put the chocolate into location 2 (as if it had never happened – contrary to facts), however, children who fail on the false belief task don't ignore or override this information which makes them answer with the chocolate's current location (also known as the realist error, Mitchell, 1994).

Perner, Sprung, and Steinkogler (2004) raised doubts that modified derivation is the critical factor enabling success on counterfactual and false belief tasks. They used two kinds of travel scenarios: (1) simple: each departure point was connected to only one destination (2) complex: each departure point was connected to both destinations. While children found it easier to answer counterfactual questions in the simple travel scenario compared to the complex one, performance on false belief questions was similar in both scenarios (in accordance with what has been found by Wellman et al., 2001). The good performance on the counterfactual task of the simple scenario ( $3\frac{1}{2}$  year olds: 72% correct) suggests that the majority of the youngest children was able to apply modified derivation but leaves open the question why this ability did not guarantee positive performance on the false belief task (0% correct).

The claim that modified derivation is the main element that makes counterfactual reasoning as well as reasoning with false beliefs difficult for children may, after all, be correct if one focuses on a different aspect of it. Peterson and Riggs focused on children's ability to ignore or override information about what is true. This ability, apparently, develops very early and is not the stumbling block for children's mastery of false belief tasks as shown by Perner

et al. (2004) and also suggested by findings on pretend play (Harris & Kavanaugh, 1993) and future hypothetical thinking (Robinson & Beck, 2000). The other aspect of modified derivation is that the modification needs to be limited to what is stated counterfactually and its logical consequences. All other propositions should remain untouched, as required by the “nearest possible world” constraint suggested by logicians (Lewis 1973; Stalnaker 1968). Features of the real world should only be changed when they are causally dependent on the counterfactual assumption (Edgington, 2011). Recent data (Rafetseder, Cristi-Vargas, & Perner, 2010; Rafetseder & Perner, in press; Rafetseder, Schwitalla, & Perner, 2012) suggest that the nearest possible world constraint inherent in modified derivation is developing between 6 and 12 years.

How children reason in these tasks can be illustrated with the stories used by Rafetseder et al. (2010, study 1). In a set up with, e.g., a high and a low shelf, children are told that when mother buys sweets she puts them on the top or the bottom shelf. When one of her children finds the sweets they take them to their respective room. The tall boy can do this regardless of where mother put them, but the small girl can only get to them on the lower shelf. If mother put the sweets on the top shelf then the girl can’t reach them and they stay there. Children are very good at remembering these general facts. They are also quite good at answering counterfactual questions in some conditions, e.g., in one condition (Example 1) mother puts the sweets on the bottom shelf, the boy comes and brings them to his room. Children are then asked: “But what, if not the tall boy but the little girl had come looking for sweets, where would the sweets be?” 71% of the 5 year olds answer correctly that then sweets would be in the girl’s room. However, in a different condition (Example 2) children have manifest problems: Mother puts the sweets on the top shelf, the boy comes and takes them to his room. “But what, if not the tall boy but the little girl had come looking for sweets, where would the sweets be?” Almost all 5 year olds give the wrong answer “in the girl’s room,” instead of “(they would have stayed) on the top shelf” (12% correct). The percentage of

children with correct answers stays low even at 9 years and then goes up to 100% correct by around 12 years (Rafetseder et al., 2012).

The authors argued that the response pattern of the younger children is due to applying ‘basic conditional reasoning’ (BCR) in response to the counterfactual question instead of counterfactual reasoning (CFR) that obeys the “nearest possible world constraint.” The younger children, when asked where the sweets would be if Y instead of X had come, are aware of the general facts about what tends to happen, but ignore the actual story and simply reason: If Y comes then the sweets tend to end up in Y’s room. So in Example 1 BCR leads to the same, correct answer as CFR does, “in the girl’s room.” In contrast, in Example 2 BCR leads to a different, wrong answer, “in the girl’s room,” (the correct CFR answer is “on the top shelf”).

To date there has been no study investigating the relationship between counterfactual reasoning and reasoning with false beliefs when also controlling for answers based on BCR. In the current study we adapted this type of stories to allow for asking a false belief question (e.g., “Where does mother think the sweets are?”) as well as the counterfactual question (“Where would the sweets be?”), in order to assess, whether counterfactual reasoning is indeed a prerequisite for reasoning with false beliefs.

### Experiment 1

For this study we use two of the conditions by Rafetseder et al. (2010) and modify them so that a false belief question can be implemented together with a counterfactual question. For instance, in the condition corresponding to Example 2 above, mum puts the sweets on the top shelf and goes to the living room to iron. From the corner of her eyes she can still see the entrance to the kitchen. Then Simon comes, ducks on his way into the kitchen (in order not to be seen by his mum), goes to the cupboard, takes the sweets from the top shelf and leaves the kitchen ducking again, and brings them to his room. Mum sees somebody tiny coming out of the kitchen and mistakenly thinks that little Julia has been searching for sweets.

Children are then asked who mum thinks was in the kitchen (little Julia) and who really was in the kitchen (Simon). They are then asked the counterfactual question “But what, if not Simon, the tall boy, but Julia, the little girl, had come along looking for sweets, where would the sweets be?” and the false belief question “Where does mum think, that the sweets are now?”. The correct answer to both questions is “on the top shelf”.

Children who use CFR should answer the counterfactual question of this condition correctly, while those that use BCR will answer wrongly “in Julia’s room.” The question of interest is now how these children will answer the false belief question. On the assumption that CFR is a necessary condition for solving the false belief task, only children who give the correct counterfactual answer should also answer the belief question correctly. Moreover, if failure to reason counterfactually is what keeps children from reasoning with false beliefs, then answers to the two questions should also strongly correlate.

On the assumption that false belief reasoning is not dependent on counterfactual reasoning, e.g., it is an independently developing domain specific module, no correlation will be expected. Presumably, children as young as 4 or 5 years, the age at which they become able to answer typical false belief questions correctly (Wellman et al., 2001), may also answer this belief question correctly, while the results by Rafetseder et al. suggest that they will answer the counterfactual question wrongly until they are older.

### *Method*

*Participants.* The study included 80 participants. The youngest group consisted of 20 children (12 males, 8 females) between 7;0 (year; month) and 8;10 ( $M = 8;1$ ,  $SD = 0;7$ ), further 20 children (10 males, 10 females) were between 9;3 and 11;0 ( $M = 10;1$ ,  $SD = 0;5$ ), and the 20 preadolescents (12 males, 8 females) were between 12;4 and 14;10 ( $M = 13;6$ ,  $SD = 0;10$ ). The control group consisted of 20 adults (3 males, 17 females) between 20;9 and 32;4 ( $M = 26;3$ ,  $SD = 2;8$ ). Participants had German as their first language and were recruited from a medium-sized city as well as from rural areas in Austria.

*Materials.* We used two different wooden models which were built on 42cm x 30cm large platforms made of wood. The model for the *sweets-story* comprised a cupboard with a shelf centrally placed and two boxes, a brown one standing on the shelf and an orange one standing beneath the shelf. For acting out the story we used sweets and three dolls, two female ones (a tall one representing the mother and a tiny one representing the daughter) and a male one twice as tall as the tiny female one (representing the son). The daughter and the son had their own rooms including a table with their photo standing on it.

The model for the *squirrel-story* consisted of a hut made of stone. A big walnut-tree was standing behind the hut. For acting out the story we used walnuts, a dwarf (who had a huge red bag for collecting nuts) and a squirrel (who had a red scarf around its neck). The dwarf lived in the dwarf village (three houses were drawn on cardboard) and the squirrel lived in a tree (including a nest).

*Design.* All children were tested in one session, which lasted about 20 min. Both story worlds (*sweets-* and *squirrel-story*) were presented and children were asked seven control questions per story world. Every child got four event sequences (two per story world). For each event sequence further seven control questions and two test questions were asked: a false belief question (e.g., “Where does mum think, that the sweets are now?”) and a counterfactual question (e.g., “But what, if not the tall boy but the little girl had come along looking for sweets, where would the sweets be?”). This means, each child got four false belief questions and four counterfactual questions in total: two in which the story-character could not have reached the item (BCR leading to the wrong answer—later referred to as condition 1) and two in which the story character could have reached the item (BCR leading to the correct answer—later referred to as condition 2). The sequence of the conditions (1 then 2 or vice versa) and the sequence of the test questions (counterfactual then false belief or vice versa) was counterbalanced between participants resulting in four different versions: (a) condition 1



first: FB then CF, (b) condition 1 first: CF then FB, (c) condition 2 first: FB then CF, and (d) condition 2 first: CF then FB.

*Procedure.* First children are familiarised with the story-world. In the sweets-story they are told that Julia and Simon love sweets and that their mum puts the sweets either in a box on the top or in a box on the bottom shelf of a cupboard in the kitchen. They are shown that Julia is not tall enough to reach the top shelf but she can reach the bottom shelf. When she finds sweets she brings them into her room. Simon, in contrast, can reach both shelves and when he finds sweets he brings them to his room. Participants are asked seven control questions: (1) Which one is the boy's room? (2) Which one is the girl's room? (3) From which shelf can the boy take sweets? (4) To where does he bring the sweets? (5) From which shelf can the girl take sweets? (6) To where does she bring the sweets? (7) Why is the girl not able to take the sweets from the top shelf?

After familiarisation with the structure of the story world testing started with two different events: in condition 1 the sweets are on the top shelf and the boy comes in search of them. In condition 2 the sweets are on the bottom shelf and the girl comes in search of them. For instance (Condition 1): "Today mum has bought some sweets and she puts them on the top shelf. The bottom shelf is empty". At this point, children are asked three control questions: (8) Where did mum put the sweets? (9) Would Simon be able to get the sweets from the top shelf? (10) Would Julia be able to get the sweets from the top shelf? Next children are told that mum has to do some ironing in the living room.

Then they watch Simon coming in search for sweets. He ducks when he enters the kitchen (in order not to be seen by his mum), goes to the cupboard, looks into the two boxes and finds sweets in the top box. When he leaves the kitchen he ducks again. But this time mum sees from the corner of her eyes that somebody tiny has come out of the kitchen. She thinks that Julia has been searching for sweets. Finally, children see how Simon brings the sweets to his room. Children are asked: (11) Who does mum think, has been in the kitchen,

searching for sweets? and (12) Who was really in the kitchen? Then children are asked the following questions (with the order of the FB and the CF question counterbalanced):

- FALSE BELIEF: “Where does mum think, that the sweets are now?”
- MEMORY 1: “On which shelf has mum put the sweets at the beginning?”
- MEMORY 2: “Where are the sweets now?”
- COUNTERFACTUAL: “But what, if not the tall boy but the little girl had come along looking for sweets, where would the sweets be?”

Wrong answers were not corrected. After answering these questions children saw a second event happening, i.e., condition 2 (the order counterbalanced between participants).

Subsequently children were familiarised with the second story which was similar in length and structure.

### *Results*

*Control and Memory Questions.* Participants answered all control questions and almost all (99%) memory questions correctly. It can therefore be assumed that problems with answering the test questions cannot be accounted for by misunderstanding the rules of the story-worlds or forgetting what has happened.

*Comparability of the Stories.* The number of correct answers to the false belief questions of the sweets story (76% correct) and the squirrel story (76% correct) did not differ. A similar result was found for counterfactual questions. Performance did not differ between the sweets story (80% correct) and the squirrel story (79% correct). Further analyses were therefore run with the number of correct answers averaged.

*Order of the Test Questions.* The false belief question was answered as often correctly whether it was asked before (79% correct) or after the counterfactual question (73% correct). In contrast, more counterfactual questions were answered correctly when asked after the false belief question (88% correct) than when asked before (71% correct), Mann-Whitney Test:  $U = 485$ ,  $z = -3.26$ ,  $p = .001$ , effect size estimate  $r = .36$ . This is surprising since Perner et al.

(2004) found a marginally significant order effect in the opposite direction, with counterfactual questions being more accurately answered when asked before the false belief question.

*Test Questions.* Figure 1 shows performance on the two test questions separately for condition 1 (BCR leading to the wrong answer) and condition 2 (BCR leading to the correct answer).

-----  
 Insert Figure 1 about here  
 -----

As expected, participants found the test questions of condition 2 (FB: 96% correct, CF: 91% correct) easier to answer than the test questions of condition 1 (FB: 56% correct, CF: 68% correct), Wilcoxon signed ranks test for FB questions:  $z = -5.62, p < .001$ , effect size estimate  $r = .63$  and for CF questions:  $z = -3.99, p < .001$ , effect size estimate  $r = .45$ . The significant difference between counterfactual questions of condition 1 and condition 2 replicates what has been found by Rafetseder et al. (2010; 2012) and supports their claim that young children answer counterfactual questions by applying BCR (leading to the correct answer in condition 2 but to the wrong answer in condition 1).

Performance between counterfactual and false belief questions was significantly correlated within conditions but, due to ceiling effects in condition 2, we only report the contingencies and correlations for condition 1: Spearman's  $r = .58, p < .001$ . Condition 1 is especially interesting, because we can distinguish children who used CFR from children who applied BCR. The high correlation between the counterfactual and the false belief task suggests that children used the same reasoning strategy on both tasks. The pattern of strategies applied on both questions of condition 1 is shown in Table 1.

-----  
 Insert Table 1 about here

-----

It can be seen that most of the participants ( $n = 44$ ) answered either all test questions of condition 1 correct or all wrong. However, 23 participants answered more counterfactual question than false belief questions correctly and only 7 participants showed the opposite pattern. Thus, it was more likely that they had problems with the false belief question than with the counterfactual question. Table 2 shows how consistent people used reasoning strategies (CFR or BCR) on counterfactual and false belief questions of condition 1. In more than two-thirds of the tasks (70%) participants used the same reasoning strategy (49% CFR, 21% BCR) to answer the counterfactual and the false-belief question. The remaining 30% were largely due to children using CFR to answer the counterfactual question but not yet the false belief question (16%), while few (5%) showed signs of the opposite combination of strategies (McNemar  $\chi^2 = 5.76, p = .03$ ), which falls within the range of the overall error rate of 9% unclassifiable responses. Overall, this result is compatible with the finding of Riggs et al. (1998) that counterfactual reasoning is a prerequisite for reasoning with false beliefs.

-----

Insert Table 2 about here

-----

### *Discussion*

The aim of study 1 was to explore whether performance on the counterfactual and on the false belief tasks is still correlated when controlling for answers based on BCR. The main result is that the majority of participants uses the same reasoning strategy for both tasks. When they use CFR for the counterfactual tasks they do so for the false belief task and when they use BCR for one task they do so for the other task. Only a minority applies different strategies, mainly of the pattern using CFR on the counterfactual task and BCR on the false belief task. This reinforces the idea that counterfactual reasoning is needed in order to reason with false beliefs.

However, one worry is that even some adults struggled to answer correctly on the false belief question of condition 1 which suggests that the stories were unnecessarily difficult. For Experiment 2 we, therefore, generally simplified the stories of Experiment 1. First of all, the cookie-story (former sweets-story) was reduced to contain only one shelf (instead of two) where mum puts the sweets. The second simplification was that the tall boy was the only child who was searching for sweets. The little girl was replaced by a dog that was searching for cookies (and the sweets were replaced by cookies to make sure that the dog could smell them). Third mum's false belief about where the cookies are were this time not based on her having a false belief about who was searching for the cookies but on an unexpected event, i.e., she put the cookies on the shelf but as soon as she was gone they fell on the floor. These simplifications were implemented accordingly in the squirrel-story too.

### Experiment 2

For the second experiment we changed the two stories of Experiment 1 slightly. In the cookie-story there was only one child (the tall boy) searching for cookies. Mum puts cookies on the shelf but as soon as she is away the cookies fall down on the floor. The dog finds the cookies on the floor and burrows them in the garden. When the boy comes home, he walks into the kitchen in search for cookies but he can't find any and leaves the kitchen. Mum watches him coming out of the kitchen and walking into his room. Children are then asked a false belief question ("Mum did not see the cookies falling on the floor. Where does she think, that the cookies are now?") and a counterfactual question ("If the cookies hadn't fallen down from the shelf, where would they be?").

Participants who reason counterfactually (CFR) are supposed to answer both questions with "in the boy's room". However, participants who do not apply CFR but BCR instead would answer with "on the shelf" because (people think that) things stay usually where they have been put. The aim of Experiment 2 is to rule out that the main result of Experiment 1 is due to a mere confusion in the stories. If CFR is in fact a necessary ingredient we would

expect the relationship between answering the counterfactual question and the false belief question correctly shown in Experiment 1 to be replicated with the simplified version of the stories in Experiment 2.

### *Method*

*Participants.* Thirty participants aged between 7;9 and 12;10 ( $M = 9;11$ ,  $SD = 1;8$ ) were tested. For later analyses they were divided into two groups. The younger group consisted of 14 children between 7;9 and 8;11 ( $M = 8;4$ ,  $SD = 0;4$ ) and the 16 participants of the older group were between 9;7 and 12;10 ( $M = 12;3$ ,  $SD = 1;1$ ). All participants had German as their first language and were recruited from rural areas in Austria.

*Design.* Each participant was tested in one session lasting about 10 minutes in a quiet area away from other children. They were given both stories but this time only condition 1. In total participants had to answer two false-belief and two counterfactual questions. The order of the stories (cookie and squirrel) and the order of the test-questions (false belief and counterfactual) were counterbalanced.

*Procedure.* In the cookie-story participants are shown that mum puts some cookies on the shelf. They are told that usually her son, Simon, comes and takes them to his room. But as soon as mum is gone the cookies fall on the floor. The dog smells the cookies, comes into the kitchen and takes them into the garden where he burrows them. Later on Simon comes in search for cookies but he cannot find any. His mum sees him coming out of the kitchen and walking back into his room. Then children are asked a control question: “Where are the cookies now?”. After they correctly answer this question they are given the two test questions:

FALSE BELIEF: “Mum did not see the cookies falling on the floor. Where does she think, that the cookies are now?”

COUNTERFACTUAL: “If the cookies hadn’t fallen down from the shelf where would they be?”

In the squirrel-story participants see that a farmer puts some nuts into a basket which is hidden in a hut (*resembling the top shelf in the cookie-story*). They are told that usually a squirrel (*resembling the boy*) squeezes through a hole in the hut and takes the nuts to a close tree trunk. But as soon as the farmer is gone the basket falls on the floor (*resembling the cookie falling on the floor*) and the nuts fall out of the basket. A boy (*resembling the dog*) who is playing close to the hut, hears the noise, has a look into the hut, sees the nuts on the floor and takes them home. Later on the squirrel comes in search for the nuts but it cannot find any. The farmer watches the squirrel coming out of the hut and running to the tree trunk. Children are asked where the nuts are now. After answering correctly “at the boy’s home” they are asked (in a counterbalanced order) the false belief question (“The farmer did not see the nuts falling out of the basket. Where does he think, that the nuts are now?”) and the counterfactual question (“If the nuts hadn’t fallen out of the basket where would they be?”). Both questions should be answered with “in tree trunk”.

## Results

*Control Questions, Comparability and Order of Stories, Order of Test Questions.* Participants answered all control questions correctly. Moreover, they answered 70% of the false belief questions of the cookie-story and 73% of the false belief questions of the squirrel-story correctly. The amount of correctly answered counterfactual questions did not differ with 80% correct answers in each story. Thus, the stories were comparable in complexity and results are therefore reported for number of correct answers averaged. Whether the cookie-story was presented before or after the squirrel-story did neither have an effect on how many counterfactual questions have been answered correctly (Mann-Whitney Test:  $U = 107.5$ ,  $z = -.25$ ,  $p = .80$ , effect size estimate  $r = .05$ ) nor on how many false belief questions have been answered correctly (Mann-Whitney Test:  $U = 106$ ,  $z = -.29$ ,  $p = .77$ , effect size estimate  $r = .05$ ). The order of the test questions had this time no effect, neither on the amount of correctly answered counterfactual questions (*for both stories the same*: Mann-Whitney Test:  $U = 97.5$ ,

$z = -.90$ ,  $p = .37$ , effect size estimate  $r = .16$ ) nor on the amount of correctly answered false belief questions (*cookie-story*: Mann-Whitney Test:  $U = 105$ ,  $z = -.39$ ,  $p = .07$ , effect size estimate  $r = .16$ ; *squirrel-story*: Mann-Whitney Test:  $U = 82.5$ ,  $z = -1.62$ ,  $p = .10$ , effect size estimate  $r = .30$ ).

*Test Questions.* In total, participants answered 80% of the counterfactual questions and 72% of the false belief questions correctly. This difference is—in contrast to study 1—not significant, Wilcoxon signed ranks test:  $z = -1.16$ ,  $p = .25$ , effect size estimate  $r = .21$ . However, performance on both test questions was—as in study 1—significantly correlated, Spearman's  $r = .58$ ,  $p = .001$ . Table 2 shows that participants applied the same reasoning strategy on both questions in 80% of the cases (68% CFR, 12% BCR). If they did not use the same reasoning process they mainly applied CFR on the counterfactual question and BCR on the false belief question (12%) and rarely the other way around (5%) which is very similar to the results of Study 1 (16% vs. 5%). Again most of the participants ( $n = 21$ ) had either all four questions correct or all four questions wrong (see Table 1). As in study 1 a few participants ( $n = 6$ ) answered more counterfactual question correctly than false belief questions compared to two participants who showed the opposite pattern. This again supports the idea that counterfactual reasoning is a necessary precondition for being able to answer the false-belief question correctly.

Figure 2 shows participants' performance on the test questions separately for age groups. The older participants answered significantly more counterfactual questions (Mann-Whitney Test:  $U = 72.0$ ,  $z = -2.25$ ,  $p = .02$ , effect size estimate  $r = .41$ ) as well as the false belief questions (Mann-Whitney Test:  $U = 65.5$ ,  $z = -2.26$ ,  $p = .02$ , effect size estimate  $r = .41$ ) correctly compared to the younger children. One aim of study 2 was to make the stories less complex. Figure 1 shows the comparison of performances between the two studies. It can be seen that the stories of study 2 were easier. However, most importantly this did not have an influence on the correlation between the two test questions.



-----  
 Insert Figure 2 about here  
 -----

### *Discussion*

The aim of study 2 was to simplify the structure of the stories and to check whether these changes influence the correlation between the number of correct answers given on the counterfactual and on the false-belief question. In fact children answered more counterfactual and false belief questions correctly in study 2 than in study 1. Thus the simplifications clearly had an effect. However, this did not change the correlation between the two test questions (the correlation being .58 in both studies). Interestingly, if children did not apply CFR it was rather the false-belief question than the counterfactual question. **This is an indication that children have to be able to reason counterfactually in order to also be able to solve false belief tasks.**

### *General Discussion*

In the introduction we noted that similar reasoning processes (modified derivation) might be involved in counterfactual and false belief tasks. However, we are claiming that the process is not about ignoring factual events such as where the chocolate has originally been put by Maxi's mum but about finding the possible world that is closest to the actual one. This claim is based on the philosophical understanding of counterfactual reasoning that one has to take the actual course of events and make only those changes to it that follow necessarily from the counterfactual assumption (Edgington, 2011; Lewis, 1973; Stalnaker, 1968). However, studies that found a link between counterfactual reasoning and false belief did not control for this nearest possible world constraint. This means that children's answers to false-belief questions could have been based on more basic if-then conclusions (such as basic conditional reasoning, BCR), which in many cases lead to the same answer as counterfactual reasoning does making it look as if children applied counterfactual reasoning.

Consistent with previous findings (German & Nichols, 2003, long chain inference; Guajardo, Parker, & Turley-Ames, 2009; Guajardo & Turley-Ames, 2004; Müller, Miller, Michalczyk, & Karapinka, 2007; Perner et al., 2004; Peterson & Bowler, 2000; Riggs et al., 1998) we found a high correlation ( $r = .58$ ) between the ability to pass a counterfactual task and a false belief task even when controlling for answers based on BCR. The majority of the children applied either CFR on both tasks or BCR on both tasks. The remaining children mostly used CFR to answer the counterfactual question and BCR to answer the false belief question. Only a minority showed the opposite answer-pattern or an unclassifiable answer-pattern. This supports previous findings that counterfactual reasoning (or its BCR substitute if that provides the same answer) is a necessary precondition of false-belief understanding (German & Nichols, 2003, short chains; Grant, Riggs, & Boucher, 2004; Müller et al., 2007; Perner et al., 2004; Peterson & Bowler, 2000; Riggs et al., 1998).

The finding that false-belief reasoning is dependent on counterfactual reasoning has important implications. It taps into the discussion about whether theory of mind is a module (ToMM) that develops in the first two years of life (Leslie, 2005) or whether it is the result of an interaction between genetics and environment (Ruffman & Perner, 2005). The present data support the latter position as children were able to reason about false beliefs only when they had developed counterfactual reasoning first. This, of course, conflicts with many previous findings that children are able to understand false beliefs when they are two years old (Southgate, Senju, & Cisbra, 2007) or even younger (Buttelmann, Carpenter, & Tomasello, 2009; Onishi & Baillargeon, 2005). All these studies show that infants understand more about the dependence of behaviour on belief than originally thought but it is not quite clear yet what it is that they understand. Ruffman and Perner (2005) pointed out that infants might not understand that other individuals hold false beliefs but predict and interpret behaviour on the basis of behaviour rules (Povinelli & Vonk, 2004), e.g., that people look for objects where they last saw them.

In favour of the innateness of theory of mind, in contrast, are findings that children with autism spectrum condition (ASC) are lacking theory of mind indicated by failing on the false belief task (Baron-Cohen, Leslie, & Frith, 1985). The idea is that the Theory of Mind Mechanism (ToMM) which represents mental states is impaired and therefore causes a lack in social understanding (Leslie, 2005). Because of these findings and the findings that counterfactual reasoning is necessary but not sufficient for belief-reasoning (Riggs et al., 1998), Peterson and Bowler (2000) studied normally developed children (ND), children with autism spectrum condition (ASC) and children who suffer from severe learning difficulties (SLD). Each child got two stories based on Riggs et al., one ending with a counterfactual question (If Mary had not baked a cake, where would the chocolate be?) and one ending with a false belief question (Where does John think the chocolate is?). While the three groups differed significantly in terms of correctly answered false belief questions (the ASC group answered significantly fewer false belief questions correctly than the ND group but there was no difference between the ASC group and the SLD group), they did not differ in terms of correctly answered counterfactual questions. Most interestingly, all children of the two clinical groups who failed on the counterfactual task also failed on the false belief task. However, from the children who passed the counterfactual task, only 60% in the SLD group and 44% in the ASC group also passed the false belief task. This shows that even in children with ASC and SLD it is the case that mastery of the false-belief task depends on giving correct answers to the counterfactual question.

It is still an open question what kinds of understanding children need to develop. According to Stenning and van Lambalgen (2008) both tasks require a possibly complex inference about a minimal change in an assumption. If children fail on the counterfactual task it is expected that they fail on the false belief task because it is assumed that in both cases there is a prepotent response that is operative. Moreover, both tasks require executive handling of interference between alternatives. However, what makes the false-belief task even

harder is the fact that children have to understand that other people's representations can be false. They have to understand the relevance of Maxi not witnessing the transfer of his object and they have to generate his false belief while for the counterfactual task the false statement is explicitly given. This clearly is an important factor in younger children's counterfactual and false belief reasoning. However, children in our studies are able to inhibit prepotent responses but still fail on some of our counterfactual and false belief tasks.

We can only bring all these findings in line with each other by arguing that children need to understand two different facets of the false belief task. In the easier task used by Riggs et al. as well as by Peterson and Bowler children need to ask themselves how the world looks from the perspective of another person. From Maxi's point of view the chocolate is still in the table drawer (location 1) while from their point of view the chocolate is in the cupboard (location 2). They need to build a world that is counter to facts, i.e., a world in which the chocolate is in location 1 (and thus inhibit prepotent responses). This, however, does not mean necessarily that children apply counterfactual reasoning (by making only those changes to the actual world that follow necessarily from the counterfactual assumption). It could be that they apply simple rules in imagined worlds. One example of this are behaviour rules which were supposed to be applied on false belief tasks (Ruffman & Perner, 2005), e.g., that people look for objects where they last saw them (for details see also Perner, 2011). Thus, children think of a world that is counter to fact (chocolate is in location 1) and then apply rules in this imagined world. That children are able to think of worlds counter to fact from very early on has been shown by children engaging in pretend play (Harris & Kavanaugh, 1993) or in future hypothetical reasoning (Robinson & Beck, 2000) and by them being able to keep several pretend play games separate from each other (Skolnick Weisberg & Bloom, 2009).

However, for the more difficult false belief tasks for which counterfactual reasoning (CFR) seems to be a necessary ingredient it is not enough to apply basic rules in imagined

worlds. Krzyzanowska (2011) argues that what is going on is similar to what has been suggested by the “Ramsey-Test”: in order to know where Maxi will search for his chocolate one has to add his perspective (the chocolate is in location 1) hypothetically to the stock of knowledge and then make minimal changes in order to maintain consistency with the assumption. In the Maxi-task, however, one cannot decide whether children in fact applied counterfactual reasoning or only basic rules, as both lead them to the same and correct answer. In our tasks, however, it becomes clear that counterfactual reasoning has to be developed in order to be also able to reason with false beliefs. Thus, “counterfactual reasoning is an integral part of belief-desire reasoning when a deviant perspective is involved” (Perner, 2011, p. ...).

### **Acknowledgements**

We would like to express our appreciation to the heads and children of the following institutions for their cooperation and valuable time participating in this project: altersgemischter Schulkindergarten Bergheim, Gemeindehort Anif, Musikhauptschule Neufelden, Privathort "San Helios" Hallein, PSV Volleyballgemeinschaft, Sportkindergarten und Krabbelstude Gröllstraße, and Volksschule Kirchberg ob der Donau. This research was financially supported by Austrian Science Fund project (I140-G15) "Counterfactual reasoning in children" as part of the ESF EUROCORES LogiCCC initiative.

## References

- Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a “theory of mind”? *Cognition*, 21(1), 37–46.
- Buttelmann, D., Carpenter, M., & Tomasello, M. (2009). Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition*, 112(2), 337–342.
- Edgington, D. (2011). Causation First: Why Causation is Prior to Counterfactuals. In C. Hoerl, T. McCormack, & S. R. Beck (Eds.), *Understanding Counterfactuals, Understanding Causation. Issues in Philosophy and Psychology* (pp. 230–241). Oxford: Oxford University Press.
- German, T. P., & Nichols, S. (2003). Children's counterfactual inferences about long and short causal chains. *Developmental Science*, 6(5), 514–523.
- 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(2), 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(Pt3), 681–702.
- Guajardo, N. R., & Turley-Ames, K. (2004). Preschoolers' generation of different types of counterfactual statements and theory of mind understanding. *Cognitive Development*, 19(1), 53–80.
- Harris, P. L., & Kavanaugh, R. D. (1993). Young children's understanding of pretense. *Monographs of the Society for Research in Child Development*, 58(1), 1–107.
- Leslie, A. M. (2005). Developmental parallels in understanding minds and bodies. *Trends in Cognitive Sciences*, 9(10), 459–462.
- Lewis, D. (1973). *Counterfactuals*. Oxford: Blackwell.

- Mitchell, P. (1994). Realism and early conception of mind: A synthesis of phylogenetic and ontogenetic issues. In C. Lewis & P. Michell (Eds.), *Children's early understanding of mind. Origins and Development* (pp. 19–45). Hove, UK: Lawrence Erlbaum Associates Ltd.
- 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(4), 615–632.
- Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, 308(5719), 255–258.
- Perner, J. (2000). About+Belief+Counterfactual. In P. Mitchell & K. J. Riggs (Eds.), *Children's reasoning and the mind* (pp. 367–397). Hove, East Sussex: Psychology Press.
- Perner, J. (2011). Theory of mind—an unintelligent design: From behaviour to teleology and perspective. In A. M. Leslie, & T. C. German (eds.), *Handbook of Theory of Mind* (pp. ....), Piscataway, NJ: Lawrence Erlbaum Associates.
- Perner, J., Sprung, M. & Steinkogler, B. (2004). Counterfactual conditionals and false belief: A developmental dissociation. *Cognitive Development*, 19(2), 179–202.
- Peterson, D. M., & Bowler, D. M. (2000). Counterfactual reasoning and false belief understanding in children with autism. *Autism*, 4(4), 391–405.
- Peterson, D. M. & Riggs, K. J. (1999). Adaptive modelling and mindreading. *Mind and Language*, 14(1), 80–112.
- Povinelli, D. J., & Vonk, J. (2004). We don't need a microscope to explore the chimpanzee's mind. *Mind & Language*, 19, 1-28.
- Rafetseder, E., Cristi-Vargas, R., & Perner, J. (2010). Counterfactual reasoning: Developing a sense of “nearest possible world”. *Child Development*, 81(1), 376–389.



- Rafetseder, E. & Perner, J. (in press). When the Alternative had been better: Counterfactual Reasoning and the Emergence of Regret. *Cognition and Emotion*.
- Rafetseder, E., Schwitalla, M., & Perner, J. (2012). *Counterfactual Reasoning: From childhood to adulthood*. Manuscript submitted for publication.
- 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(1), 73–90.
- 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, East Sussex: Psychology Press.
- Ruffman, T., & Perner, J. (2005). Do infants really understand false belief? *Trends in Cognitive Sciences*, 9(10), 462–463.
- Skolnick Weisberg, D., & Bloom, P. (2009). Young children separate multiple pretend worlds. *Developmental Science*, 12(5), 699–705.
- Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science*, 18(7), 587–592.
- Stalnaker, R. (1968). A theory of conditionals. *Studies in Logical Theory*, 2, 98–112.
- Stenning, K., & Van Lambalgen, M. (2008). *Human reasoning and cognitive science*. Cambridge: MIT Press.
- Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: the truth about false belief. *Child Development*, 72(3), 655–684.
- Wimmer, H. & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13(1), 103–128.

## Table Captions

*Table1.* Number of participants responding to the two CF and the two FB Questions of Condition 1 in Study 1 and 2

*Table2.* Amount of Reasoning Strategies (% in parentheses) applied on FB and CF Questions of Condition 1 in Study 1 and 2

Table 1

		FB Question			
		++	+-	-+	--
Study 1					
CF Question	++	30	2	8	5
	+-		4		5
	-+	3	1	2	3
	--	3			14
Study 2					
CF Question	++	18		3	2
	+-				
	-+			1	1
	--	1	1		3

Table 2

		FB Question		
		CFR	BCR	Other
Study 1				
CF Question	CFR	79 (49%)	25 (16%)	4 (2%)
	BCR	8 (5%)	33 (21%)	5 (3%)
	Other	2 (1%)	3 (2%)	1 (1%)
Study 2				
CF Question	CFR	41 (68%)	7 (12%)	1 (2%)
	BCR	3 (5%)	7 (12%)	2 (3%)

### Figure Captions

*Figure1.* Correct answers (%) given to the counterfactual and the false belief questions in condition 1 and condition 2 of study 1

*Figure2.* Correct answers (%) given to the counterfactual questions in condition 1 of study 1 and study 2

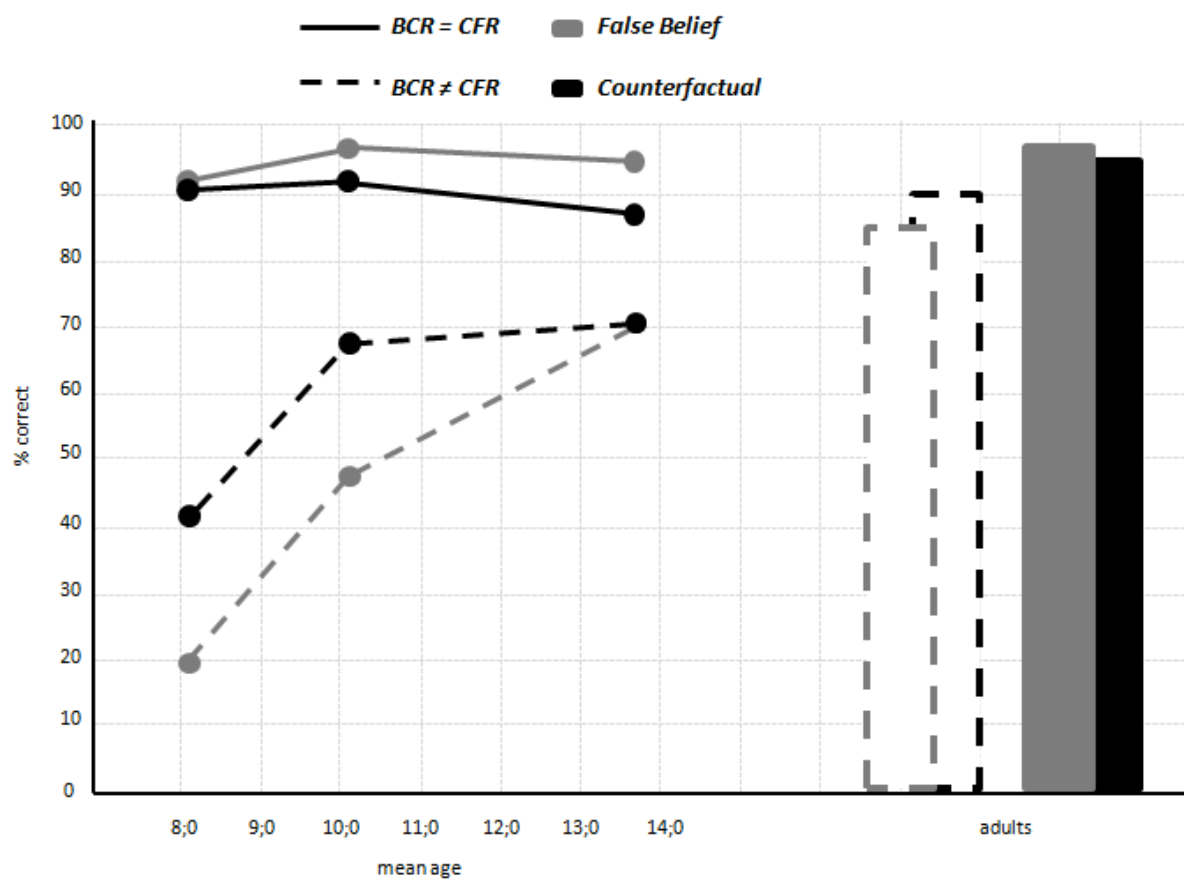


Figure 1.

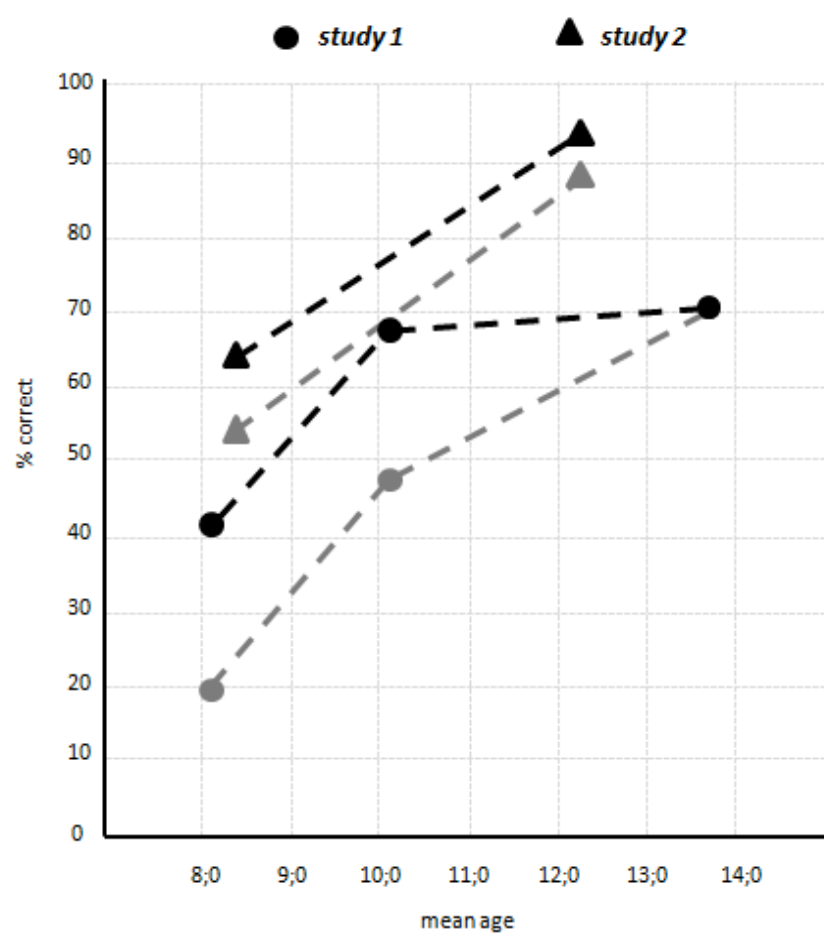


Figure2.