Journal of Experimental Child Psychology 114 (2013) 389–404

Contents lists available at SciVerse ScienceDirect

Journal of Experimental Child Psychology

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

Counterfactual reasoning: From childhood to adulthood

Eva Rafetseder [a,*], Maria Schwitalla [a], Josef Perner [a,b]

[a] *Department of Psychology, University of Salzburg, 5020 Salzburg, Austria*
[b] *Centre for Neurocognitive Research, University of Salzburg, 5020 Salzburg, Austria*

ARTICLE INFO

Article history:
Received 16 February 2012
Revised 23 October 2012
Available online 5 December 2012

Keywords:
Counterfactual reasoning
Basic conditional reasoning
Executive functioning
Childhood
Adulthood
Nearest possible world constraint

ABSTRACT

The objective of this study was to describe the developmental progression of counterfactual reasoning from childhood to adulthood. In contrast to the traditional view, it was recently reported by Rafetseder and colleagues that even a majority of 6-year-old children do not engage in counterfactual reasoning when asked counterfactual questions (*Child Development*, 2010, Vol. 81, pp. 376–389). By continuing to use the same method, the main result of the current Study 1 was that performance of the 9- to 11-year-olds was comparable to that of the 6-year-olds, whereas the 12- to 14-year-olds approximated adult performance. Study 2, using an intuitively simpler task based on Harris and colleagues (*Cognition*, 1996, Vol. 61, pp. 233–259), resulted in a similar conclusion, specifically that the ability to apply counterfactual reasoning is not fully developed in all children before 12 years of age. We conclude that children who failed our tasks seem to lack an understanding of what needs to be changed (events that are causally dependent on the counterfactual assumption) and what needs to be left unchanged and so needs to be kept as it actually happened. Alternative explanations, particularly executive functioning, are discussed in detail.

Introduction

Throughout the day, adults drift off into fictional worlds when they watch movies, read books, tell fairy tales, think about their future, or moan about missed opportunities. The ability to imagine fictional worlds can already be observed in very young children when they create imaginary companions (Taylor, 1999) or engage in pretend play (e.g., pouring pretend tea into a cup by lifting an empty teapot

* Corresponding author.
E-mail address: eva.rafetseder@sbg.ac.at (E. Rafetseder).

0022-0965
http://dx.doi.org/10.1016/j.jecp.2012.10.010

above the cup and tilting the teapot as described by Harris & Kavanaugh, 1993). Although these are very different ways of imagining, they have at least one feature in common: They are to a certain extent similar to the real world. This is because fictional worlds can never be described to a full extent, and missing information needs to be imported from real-world knowledge (see Skolnick Weisberg & Goodstein, 2009, for factors that determine what is imported).

Importing world knowledge into imagined worlds is also crucial for counterfactual reasoning (CFR). Creating a counterfactual world has been defined as creating an imagined world as close as possible to the actual world (Lewis, 1973). CFR "involves a change in some features of the actual world in addition to those required by the truth of the antecedent of the counterfactual, while other such features are left unchanged" (Woodward, 2011, p. 21). Edgington (2011) clarified that features should be changed only when they are causally dependent on the antecedent of the counterfactual. The fact that adults are quite successful in reaching a consensus about counterfactuals suggests that CFR is a highly structured process (Pearl, 2011).

Rafetseder, Cristi-Vargas, and Perner (2010) investigated whether adults and children agree on what needs to be changed and what needs to be left unchanged in order to create a counterfactual world. They used stories that were acted out with dolls. For instance, in one story world, a mother sometimes placed candy on the top shelf and sometimes placed it on the bottom shelf of a cupboard. In Studies 2 and 3, her tall son could only reach the top shelf (he had his leg in a cast and could not kneel down to reach the bottom shelf), and his little sister could only reach the bottom shelf (she was not tall enough to reach the top shelf). When the boy came looking for the candy, it ended up in his room only if it had been placed on the top shelf. If the candy had been placed on the bottom shelf, he was unable to reach it and it remained there. Similarly, when his sister came looking for the candy, it ended up in her room only if it had been placed on the bottom shelf. If the candy had been placed on the top shelf, it remained there.

Adults and children had no problem in predicting what would happen when, for example, the candy was on the top shelf and the girl came; the candy would stay on that shelf and so forth (Rafetseder et al., 2010). Adults, unlike children, were also able to answer counterfactual questions such as the following (Example 1). The mother placed the candy on the top shelf, and the boy took it to his room. What if the little girl had come instead of the boy? Where would the candy be? Remarkably, 100% of the adults, but only 24% of the 6-year-olds, answered correctly that the candy would have stayed on the top shelf. None of the adults, but most of the children, said that the candy would have ended up in the girl's room (which would be the result in a possible world where the candy was on the bottom shelf—in contrast to where it actually was—and the fact that the girl came to get it). There was a clear consensus among adults, but not among children, of what needs to be changed to be in accordance with the counterfactual assumption and what needs to be left unchanged and so needs to be imported from the real course of events.

This conclusion gives rise to other questions such as what the underlying cognitive processes are and when we are fully capable of CFR (in an adult-like sense). Existing data do not provide clear answers to these two questions. Regarding the second question, a few studies reported that 3- to 4½-year-olds are able to answer counterfactual questions correctly (Beck, Robinson, Carroll, & Apperly, 2006, standard counterfactuals; German & Nichols, 2003; Guajardo, Parker, & Turley-Ames, 2009; Guajardo & Turley-Ames, 2004; Harris, German, & Mills, 1996; Kuczaj & Daly, 1979; Riggs, Peterson, Robinson, & Mitchell, 1998). Other studies did not find this ability before 5 or 6 years of age (Beck & Crilly, 2009, open counterfactuals; Beck et al., 2006, open counterfactuals; Rafetseder et al., 2010)—especially when they tested for feelings of regret and relief (Amsel & Smalley, 2000; for further details, see Amsel et al., 2003; Beck & Crilly, 2009, regret stories; Burns, Riggs, & Beck, 2012; Guttentag & Ferrell, 2004; O'Connor, McCormack, & Feeney, 2012; Rafetseder & Perner, 2012; Weisberg & Beck, 2010; Weisberg & Beck, 2012). Anticipation of regret, which is important for avoiding negative outcomes in the future (Epstude & Roese, 2008), does not develop before 9 years of age (Guttentag & Ferrell, 2008).

To date, it is not clear what accounts for these age discrepancies. Because children's executive functioning improves throughout preschool (Carlson, 2005), it has been suggested that executive functioning explains these discrepancies.

Independently of executive functioning, Rafetseder et al. (2010) gave a further explanation for the discrepant findings of when CFR emerges: Tasks that are successfully completed early in life can be solved by a simpler reasoning strategy than those that are successfully completed later. That is, the counterfactual problems that children find to be easy can be solved by applying conditionals that express general regularities to suppositions that can be counter to fact. Rafetseder and colleagues classified this as a case of basic conditional reasoning (BCR). For instance, in Example 1 above, when the mother placed the candy on the top shelf and the boy brought it to his room, children below a certain age treated the counterfactual question "Where would the candy be if the girl had come instead of the boy to look for it?" by simply disregarding the fact that the mother had placed it on the top shelf and answered that it would be in the girl's room.[1] In other words, although they could reason with antecedents counter to fact (if the girl had come, the candy would not be where it is now, in the boy's room, but rather somewhere else), they failed to take into account the constraint on CFR to stick as closely as possible to the given facts. The fact that the mother had placed the candy on the top shelf needs to be imported from the real course of events because it is causally independent of the counterfactual antecedent (the girl comes looking for the candy), a principle that Rafetseder and colleagues referred to as the *nearest possible world constraint* (based on Lewis, 1973). All adult participants obeyed this constraint, but fewer than a quarter of the 6-year-old participants gave the correct answer in Example 1 (Rafetseder et al., 2010, 24% correct).

Interestingly, when the mother placed the candy on the bottom shelf and the boy came in search of it (Example 2), children had little difficulty in answering the counterfactual question that the candy would be in the girl's room if the girl had come instead of the boy (Rafetseder et al., 2010, 88% correct). The difference in difficulty between Example 1 and Example 2 was explained as follows. In Example 1 (candy placed on the top shelf), BCR gives a different answer (in the girl's room) than CFR (on the top shelf). On the contrary, in Example 2 (candy placed on the bottom shelf), CFR and BCR give the same answer (in the girl's room). Rafetseder and colleagues (2010) concluded that studies in which very young children are able to give the correct answer to counterfactual questions (e.g., the stories used by Harris et al., 1996) did not take into consideration children's use of BCR. To our knowledge, there is only one study that checked for answers based on BCR, and this study was conducted with children who were 6 years of age or younger and whose performance was close to floor in the critical condition (Rafetseder et al., 2010). So, our first aim was to establish when children become able to give the same answer as adults in this condition.

Study 1

The current study used the paradigm of Rafetseder and colleagues (2010, Studies 2 and 3), who collected data with 5- and 6-year-old children as well as with adults ranging in age from 14;7 to 75;10 (years;months). The main aim of the current study was to look at the developmental trajectory of CFR beyond 6 years of age with 9- to 14-year-olds when controlling for answers based on BCR.

Method

Participants

The sample consisted of 34 children and adolescents (11 girls and 23 boys) from a nursery school, two youth centers, and a scout group. The age range was between 9;0 and 14;5. The mean age was 11;9 with a standard deviation of 2;0. For later analysis, the overall sample was split into two age groups. The ages of the 18 children ranged from 9;0 to 11;1 (*M* = 10;0, *SD* = 0;7), and the ages of the 16 adolescents ranged from 12;5 to 14;5 (*M* = 13;9, *SD* = 0;8). In this and the subsequent study, children and adolescents were recruited by writing to the parents of the children and adolescents from

[1] Obermayr (2011) showed that when children are told that the mother has placed the candy into one of the boxes (without being told which box) and are then asked "What will happen to the candy when the girl comes looking for it?" 71% (n = 17) of the 6- to 9-year-olds answered accordingly with "it will end up in the girl's room", whereas only one child stated that more information is needed in order to be able to answer this question. Three further children went for the assumption that the girl is not tall enough and, therefore, that the candy will stay in the top box.

each institution that participated in the study. They were included only if their parents gave written consent. Participants spoke German as their first language and came from a mixed working- and middle-class population.

Materials

To control for story-specific effects, we used two different story worlds sketched with two different wooden models that were built on 42×30-cm platforms. The model for the *candy story* consisted of a cupboard with a shelf centrally placed and two boxes: a brown one placed on the shelf and an orange one placed beneath that shelf. We used candy and two dolls, a female one and a male one twice as tall as the female figure, to act out the story. Each doll had its own room, including a table and a photo of the female doll in one room and the male doll in the other room.

The model for the second story, the *dwarf story,* consisted of a hut and a large walnut tree located behind it (the hut represented the top shelf and the walnut tree represented the bottom shelf of the candy story). We used walnuts, a dwarf (representing the boy), and a squirrel (representing the girl) to act out the story. The dwarf lived in the dwarf village (three houses were drawn on a piece of cardboard, representing the boy's room) and the squirrel lived in a nest in another tree (representing the girl's room).

Design

Each child was tested individually in a 20-min session during which we presented both story worlds (candy story and dwarf story) and asked eight control questions per story world to be sure that the children remembered all of the details. For the candy story, these questions were as follows. Which one is Simon's/Julia's room? (Questions 1 and 2). From which shelf can Simon/Julia take the candy? (Questions 3 and 4). Where does Simon/Julia bring the candy to? (Questions 5 and 6). Why is Simon/Julia not able to take the candy from the bottom/top shelf? (Questions 7 and 8). For the dwarf story, we asked the same questions again, tailored to the dwarf and the squirrel.

Each child was told about and watched two events (one per story world), for example, that the candy is on the top shelf today. The child was asked an indicative future question for each event (e.g., "What will happen to the candy when the boy comes looking for it?") and a subjunctive past (counterfactual) question (e.g., "What if the little girl had come looking for the candy instead of the boy? Where would the candy be?"). Each child answered a total of two indicative future questions and two subjunctive past questions. For one subjunctive past question, BCR and CFR resulted in a different answer (as was the case for Example 1 described above), and for the other subjunctive past question, BCR and CFR resulted in the same answer (as described in Example 2 above). The order of the stories and the conditions were counterbalanced.

Procedure

Each child was tested at his or her respective institution in a quiet area away from the other children. The candy story was about Simon and his little sister Julia, both of whom had their own room (portrayed with a table and a photo of each protagonist). There was also a kitchen (portrayed with a cupboard and two boxes: one on the bottom shelf of the cupboard and one higher up). When their mother bought the candy, she placed it *either* in the box on the top shelf *or* in the box on the bottom shelf. Simon was tall enough to reach the top shelf, but he had his leg in a cast and could not kneel down to reach the bottom box. Julia could reach the bottom shelf, but she was not tall enough to reach the top shelf. If Julia found candy on the bottom shelf, she brought it into her room, and if Simon found candy on the top shelf, he brought it into his room.

The dwarf story was about a dwarf who lived in a dwarf village and a squirrel that lived in a nest on a tree. Both the dwarf and the squirrel liked nuts very much. In their search for nuts, they came across a nut tree under which a hut had been built. The hut had a hole in its roof, and the nuts fell through this hole into the hut. The dwarf was too clumsy to climb up the nut tree, but he could open the door of the hut to collect some nuts. The squirrel, in contrast, could not open the door of the hut, but it could climb up the tree to get to the nuts. Note that all of the nuts were *either* in the hut *or* on the nut tree. If the dwarf found nuts in the hut, he brought them back to the dwarf village, and if the squirrel found nuts on the nut tree, it brought them to its nest.

After familiarizing the participant with the structure of a story world, testing started with different events. Each event was a combination of where the object of desire was located (top shelf–bottom shelf, hut–nut tree) and which character came to collect it (boy–girl, dwarf–squirrel). For each event, an indicative future question and a subjunctive past question were asked. The following is an example: "Today the bottom shelf is empty. There is only candy on the top shelf." At this point, the child was asked the following questions:

Memory 1: "Where is the candy now?"
Indicative future: "What will happen to the candy when the boy comes looking for it?" (German original version: "Was passiert mit dem Zuckerl wenn jetzt der Simon kommt und nach Zuckerl sucht?")

After the child had given an answer, the indicative future event was played out: "Look! This time the boy comes looking for the candy. He finds it on the top shelf and takes it to his room!"

Memory 2: "Where is the candy now?"
Subjunctive past: "What if the little girl had come looking for the candy instead of the boy? Where would the candy be?" (German original version: "Aber, wenn nicht der Simon sondern die kleine Julia nach Zuckerl gesucht hätte, wo wäre denn dann das Zuckerl?")

Results

Control, memory, and indicative future questions

Children answered all control and memory questions correctly. Performance on the indicative future questions was also at ceiling. Children understood that under some conditions the protagonist could not reach the item.

Subjunctive past questions

Fig. 1 shows the children's performance on the subjunctive past questions of the current study (black circles) and of Studies 2 and 3 reported by Rafetseder and colleagues (2010, gray triangles). The results are divided to show the conditions (a) in which BCR resulted in a different answer than CFR (dashed lines) and (b) in which BCR resulted in the same answer as CFR (solid lines).

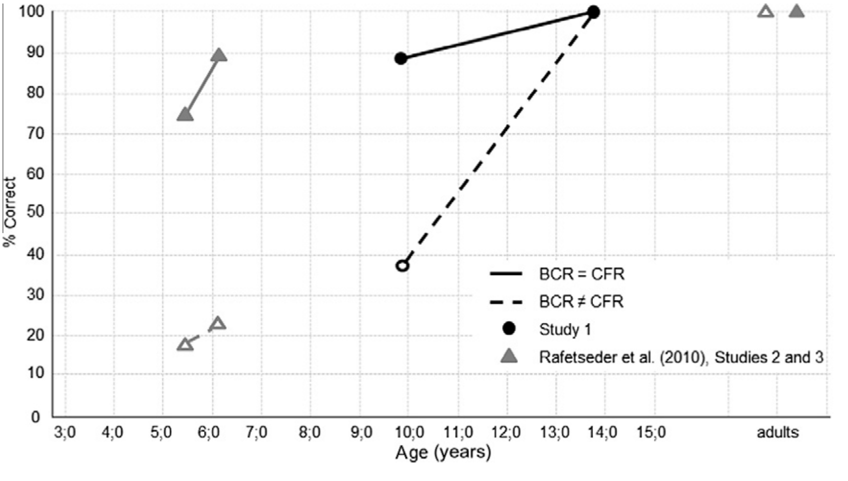

Fig. 1. Percentages of correct answers in the current Study 1 (black circles) and from Rafetseder and colleagues (2010, Studies 2 and 3) (gray triangles). Adapted with permission from "Counterfactual reasoning: Developing a sense of 'nearest possible world'" by E. Rafetseder, R. Cristi-Vargas, and J. Perner, 2010, *Child Development*, Vol. 81, p. 384.

In the most critical conditions (discriminating answers based on BCR from answers based on CFR, represented by the dashed lines in Fig. 1), the performance of the 9- to 11-year-olds was significantly different from the performance of the 12- to 14-year-olds; whereas the 9- to 11-year-olds answered only 39% of the subjunctive past questions correctly ($M = 0.39$, 99% confidence interval [*CI*] [0.05, 0.73]), the 12- to 14-year-olds answered all subjunctive past questions correctly, $U = 56.0$, $z = -3.75$, $p < .001$; effect size estimate: $r = .64$. The 9- to 11-year-olds' performance did not differ significantly from chance level, $\chi^2(2, N = 18) = 2.33$, $p = .31$. The chance level was taken to be .33, assuming that the four different locations (top, bottom, girl's room, and boy's room) were possible answers and subtracting the location where the item really was because children at this age did not commit the realist error anymore. They systematically gave the wrong answer (73% of all errors) and said that the searcher would have taken the item (even though he or she would not have been able to do so). The two age groups did not differ in terms of correct answers for the other conditions where BCR and CFR resulted in the same answer (solid line), $U = 128.0$, $z = -1.35$, $p = .18$; effect size estimate: $r = .23$.

The performance of the 9- to 11-year-olds (current study) was similar to that of the 6-year-olds in Rafetseder and colleagues (2010, Study 2). In contrast, the performance of the 12- to 14-year-olds (current study) was very different from that of the 6-year-olds in the most critical conditions (dashed line in Fig. 1 and 24% correct) but not in the other conditions (solid line, 88% correct). In fact, the performance of the 12- to 14-year-olds (current study) was similar to that of the adult control group in Rafetseder and colleagues (2010, Study 3) for all conditions.

Discussion

CFR means creating the nearest possible world (Lewis, 1973). In other words, when reasoning counterfactually, one needs to create an imagined world that is causally compatible with the counterfactual assumption (Edgington, 2011). Unknown factors in this imagined scenario, however, must be filled with facts from the real-world scenario. In contrast to other types of reasoning, CFR must coordinate the counterfactual scenario with the real-world event. It was concluded recently (Rafetseder et al., 2010) that adults, but not children, follow this constraint. Apparently, younger children do not take into account special instances of the real-world event except by default, chance, or natural bias, and they make assumptions using BCR. Study 1 investigated how CFR developmentally progresses from childhood to adulthood.

To our surprise, it was not until 12 years of age that all children responded in an adult-like manner, that is, answered correctly in those conditions where BCR led to the wrong answer (only CFR resulting in the correct answer). One concern, however, is that the story world of Study 1 is unnecessarily complicated. It would be reassuring if the same difficulties with CFR could be replicated in a simpler story world. The aim of Study 2 was to create a setting where children do not need to learn and memorize arbitrary rules.

Study 2

For Study 2, we used story worlds that could be understood without children needing to learn a set of rules at the start. For instance, children were shown a clean floor. Then they saw that Carol came home and did not take her dirty shoes off, and she made the floor all dirty with her shoes. It has been reported (Harris et al., 1996) that even 3-year-olds answered 75% of the past subjunctive questions correctly (e.g., "If Carol had taken her shoes off, would the floor be dirty or clean?"). One problem with this procedure is that with the children's answer "clean," it is impossible to determine whether they used BCR ("When somebody takes dirty shoes off, floors tend to stay clean") or CFR ("If Carol had taken her shoes off, then she would not have made the clean floor dirty and the floor would have stayed clean").

In Study 2, therefore, we introduced a second puppet, Max, who made the clean floor dirty with his shoes right after Carol did so. When asked "What if Carol had taken her shoes off? Would the floor be dirty or clean?" CFR led to a "dirty" answer (if Carol had taken her shoes off, then she would not have made the floor dirty, but the floor would have been made dirty by Max anyway). In contrast, BCR led to the answer "clean" because when people take their dirty shoes off, floors tend to stay clean, and so if Carol takes her shoes off, the floor will stay clean.

To prevent confusion, we refer to the counterfactual task that can be solved even with BCR as the undiscriminating task (BCR = CFR) and refer to the counterfactual task that can be solved only with CFR as the discriminating task (BCR ≠ CFR). We are not claiming that participants who apply CFR on both tasks would find it easier to do so on the undiscriminating task than on the discriminating task. The main difference is that we can distinguish BCR from CFR in the discriminating task, whereas this is not possible in the undiscriminating task. Based on Study 1, we expect that children up to 11 years of age find the task in which BCR leads them to the correct answer to be easy, but find the task in which BCR leads them to the wrong answer to be difficult. As children increase in age, this difference should vanish.

Method

Participants

There were 20 kindergarten children (14 girls and 6 boys) aged 5;0 to 6;1 (M = 5;8, SD = 0;4), 20 younger school children (7 girls and 13 boys) aged 7;9 to 10;8 (M = 9;9, SD = 0;11), 21 older school children (5 girls and 16 boys) aged 13;6 to 15;3 (M = 14;5, SD = 0;5), and 21 adults (14 women and 7 men) aged 22;9 to 67;8 (M = 31;0, SD = 11;10). The kindergarten children were recruited from a nursery school in a rural area, and the school children were recruited from two schools in a mid-sized city. All institutions were serving a mixed working- and middle-class population. Participants of the oldest age group were mostly students or adults with a higher level of education. All participants spoke German as their first language.

Materials

We reused the story that Harris and colleagues used in their 1996 study and developed three further stories. For each of the four stories, there was an undiscriminating version and a discriminating version, and we used the same material in both versions. For the *painting* and *sleeping* stories, we used puppets that were approximately 10 cm tall and made of wood and cloth as well as wooden plates and some essential props. For the *becoming wet* story, we used a plastic figure approximately 8 cm tall, a plastic plate with a little garden, a pool, and a spray bottle with water to simulate rain. For the *dirty shoes* story, we used plastic figures approximately 4 cm tall, a plastic plate, and a hazel nut spread to simulate dirt. For each story, we had two pictures showing the possible final states of the story. This was used as a different answer format to enhance the performance of the youngest participants who might have problems giving a verbally correct answer.

Design

Each participant was given the four different stories either in the order of (1) dirty shoes, (2) sleeping, (3) becoming wet, and (4) painting or in the reversed order (4 to 1). The stories were told either in the undiscriminating version or in the discriminating version, and each participant was told two undiscriminating and two discriminating versions. The order in which these versions were given was fully counterbalanced. To avoid answer biases, the order of the verbs of interest was completely balanced in the subjunctive past question ("... would the floor be clean or dirty?" vs. "... would the floor be dirty or clean?"). In addition, in each version, the participant needed to answer a *now control question* (for the dirty shoes task: "What does the floor look like now?") and a *before control question* ("What did the floor look like before Susi/the children walked over it?").

Procedure

The test was conducted in a quiet area at the respective institution and took approximately 15 min. The procedure was the same in each age group. Every participant was told all four stories, which were acted out with props. The participant was then asked questions that needed to be answered verbally and by also pointing to a picture. For example, in the undiscriminating version of the dirty shoes story, the background was explained as follows: "One day the floor is nice and clean, but then something happens. Susi comes home and doesn't take her shoes off. She walks in and makes the floor all dirty."

Then, two control questions were asked to check whether the story was fully understood: "Is the floor clean or dirty now?" and "Was the floor clean or dirty before Susi walked in?" After this, the subjunctive past question was asked: "What would have happened if Susi had taken her shoes off? Would the floor be clean or dirty?" (German original version: "Was wäre wenn die Susi die Schuhe ausgezogen hätte? Wäre der Boden dann sauber oder dreckig?"). After answering verbally, the participant was shown two pictures, one of a clean floor and one of a dirty floor, and was asked, "What would it look like?" In the discriminating version, the story was slightly different: "One day the floor is nice and clean, but then something happens. Susi and Max come home and they don't take their shoes off. They walk in . . .." From here onward, the story was the same as in the undiscriminating version, and the same questions were asked and the same pictures of the floor were shown: one with one pair of dirty footprints and one with no dirty footprints. It is important to note that Susi walked in first and Max walked in immediately after her. As a consequence, the floor was clean before Susi walked in, as was the case in the undiscriminating version. This was to ensure that the answers to the control questions were the same across conditions and to prevent these from having an impact on the subsequent counterfactual questions.

The logic of the other three stories was the same. In the undiscriminating version of the sleeping story, a baby was sleeping when her older sister came home, started playing the drums, and woke the baby up. Then, the child was asked whether the baby would be awake or asleep if the sister had not played the drums. In the discriminating version, the sister was again playing the drums, but also their grandma dropped a pot, which would have woken up the baby anyway. In the becoming wet story, a girl was playing in the yard when it suddenly started raining. If it had not started raining, the girl would have been dry in the undiscriminating version, but because of also jumping into a pool, the girl would still have been wet in the discriminating version. In the painting story, Hans was drawing a house on a plain sheet of paper. In the undiscriminating version, the sheet of paper would be blank if he had not started drawing. In the discriminating version, another child also drew on the paper. So, if Hans had not drawn anything, there would still be another drawing on the sheet of paper.

Results

Control questions

Answers to the now and before control questions were quite accurate. Only four 5- and 6-year-olds and one 14-year-old made mistakes. Each of these children gave only one wrong answer. For this reason, they were not excluded from the sample. In total, 99% of the now control questions and 98% of the before control questions were answered correctly. It can be assumed that children understood and remembered the content of the stories very well.

Subjunctive past questions

Overall, the four stories were comparable in difficulty (correct answers: 68% for dirty shoes, 61% for sleeping, 68% for becoming wet, and 69% for painting), Friedman's test, $\chi^2(3, N = 82) = 0.36$, $p = .31$. Therefore, we ran analyses averaged across all four stories. Participants answered significantly more subjunctive past questions correctly in the undiscriminating versions (98%, $M = 1.96$, 99% *CI* [1.91, 2.02]) than in the discriminating versions (64%, $M = 1.28$, 99% *CI* [1.03, 1.53]), $z = -5.40$, $p < .001$; effect size estimate: $r = .30$. Fig. 2 shows participants' performance on the subjunctive past questions according to age groups.

The adult group answered all subjunctive past questions in the undiscriminating version correctly and answered nearly all of them correctly in the discriminating version (95%, $M = 1.90$, 99% *CI* [1.72, 2.09]), resulting in a nonsignificant difference ($z = -1.41$, $p = .16$; effect size estimate: $r = .15$).

The 13- to 15-year-olds' performance differed significantly ($z = -2.24$, $p = .03$; effect size estimate: $r = .24$) between the undiscriminating condition (100%) and the discriminating condition (88%, $M = 1.76$, 99% *CI* [1.49, 2.03]). A McNemar test (based on a binomial distribution—children answered either one or two of two questions correctly), however, revealed a marginally significant result ($p = .06$). Focusing only on the discriminating version, the 13- to 15-year-olds performed similarly well as adults did, $U = 189.0$, $z = -1.23$, $p = .22$; effect size estimate: $r = .19$.

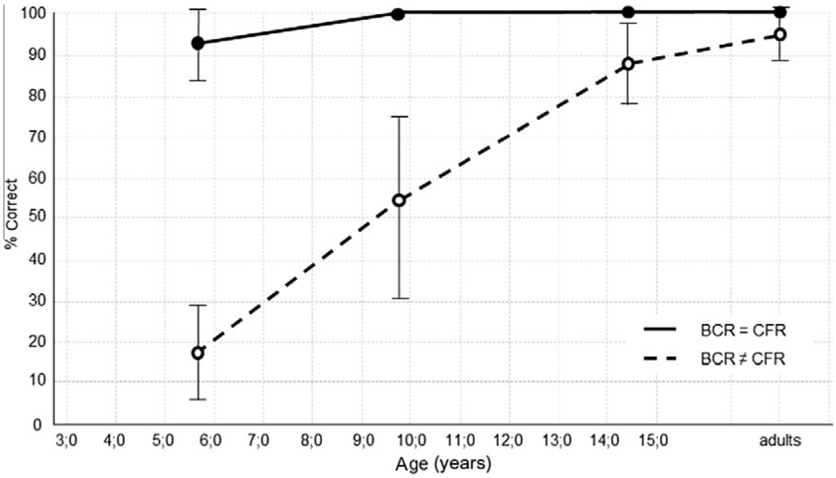

Fig. 2. Percentages of correct answers in the current Study 2. Error bars represent 95% confidence intervals.

All errors in the group of the 13- to 15-year-olds occurred exclusively in the sleeping story. A baby was sleeping when her older sister came home and started playing the drums. Subsequently, their grandma dropped a pot and the baby woke up. Five adolescents stated that if the sister had not played the drums, the baby would not have woken up. Interestingly, three of them explained that this would have been the case because their grandma would not have dropped the pot either. We counted these answers as incorrect for the following reasons. First, the counterfactual antecedent did not ask participants to make these assumptions (as also indicated by adults' answers). Second, these problems did not occur in the other stories because the two events were clearly causally independent and these stories did not differ significantly from the sleeping story in terms of correct answers.

The 7- to 10-year-olds gave perfect answers in the undiscriminating version but not in the discriminating version (53%, $M = 1.05$, 99% *CI* [0.45, 1.65]), $z = -3.07$, $p = .002$; effect size estimate: $r = .34$. The number of correct answers in the discriminating version differed significantly from a random distribution, $\chi^2(2, 20) = 9.90$, $p = .007$. Among children in this age group, 40% ($n = 8$) gave no correct answer, 15% ($n = 3$) gave one correct answer, and 45% ($n = 9$) gave two correct answers. A comparison of the 7- to 10-year-olds with the older participants with respect to the discriminating version revealed significant differences (compared with the 13- to 15-year olds: $U = 124.5$, $z = -2.56$, $p = .01$; effect size estimate: $r = .40$; compared with the adults: $U = 106.5$, $z = -3.29$, $p = .001$; effect size estimate: $r = .51$.

The 5- and 6-year-olds gave 93% correct answers ($M = 1.85$, 99% *CI* [1.62, 2.08]) in the undiscriminating version but gave only 18% ($M = 0.35$, 99% *CI* [0.04, 0.66]) in the discriminating version, $z = -3.87$, $p < .001$; effect size estimate: $r = .43$. Their numbers of correct answers in the discriminating version differed significantly from a guessing distribution (25%: no correct answers; 50%: one correct answer; 25%: two correct answers), $\chi^2(2, 20) = 18.70$, $p < .001$. Among children in this age group, 65% ($n = 13$) gave no correct answer, 35% ($n = 7$) gave one correct answer, and no child answered both subjunctive past questions of the discriminating version correctly. The 5-year-olds gave significantly fewer correct answers in the discriminating version than all of the other groups (smallest difference with the 7- to 10-year-olds: $U = 118.5$, $z = -2.42$, $p = .02$; effect size estimate: $r = .38$).

Discussion

Results in the undiscriminating counterfactual condition confirm what was reported in 1996 by Harris and colleagues with 3-year-olds who answered 75% of counterfactual questions correctly. In our study, 5- and 6-year-olds answered 93% correctly. However, the findings by Rafetseder and colleagues (2010) and in our Study 2 raise the question of whether children at this early age were giving correct answers because they were applying CFR or because they were using BCR. In Study 2, therefore, we used the discriminating version of the task by Harris and colleagues (1996), in which answers based on CFR could be distinguished from answers based on BCR. The 5-year-olds' correct answers dropped to 18%, and even the 7- to 10-year-olds answered only 53% of the questions correctly. It was not until around 13 to 15 years of age that the participants reached adult-like performance.

The results of Study 2 support the findings of Study 1, in which the 9- to 11-year-olds answered only 39% of the questions correctly in the condition where answers based on CFR differed from those based on BCR. Even though there is a tendency for children to find the task in Study 2 slightly easier than the task in Study 1, this tendency is nominal.

Study 2 also supports results that were published by Ferrell, Guttentag, and Gredlein (2009). They told children stories about two characters who suffered from the same outcome, for example, falling off a swing that was slightly broken. The difference was that Character A could have chosen another swing that was intact. So, if Character A had gone for the other swing, she would not have fallen off the swing. In contrast, if Character B had gone for the other swing, she would have fallen off anyway because the other swing was broken too. In the explicit version, children were asked whether one of these two characters felt worse than the other. Only 36% of the 8-year-olds reasoned that Character A would have felt worse (because this character could have gone for a better option). If, however, children reasoned with a simpler reasoning strategy on both tasks (i.e., that choosing a different swing—based on the counterfactual antecedent—would have resulted in a different outcome), they arrived at the same answer for both characters (not falling off the swing and so feeling okay, disregarding the fact that the second swing for Character B was broken). The majority of the 8-year-olds concluded that both characters felt the same, which is yet another indication that they did not apply CFR.

General discussion

The results of the reported studies suggest that children's ability to reason counterfactually is not fully developed in all children before 12 years of age (showing great variability in performance, especially at around 10 years). These results, together with previous work, suggest a continuum along which children gradually come to engage in adult-like counterfactual reasoning. Rafetseder and colleagues (2010) suggested that the counterfactual tasks that even young children complete with no difficulty may be solved with simpler reasoning strategies (e.g., by BCR). When checking for answers based on simpler reasoning strategies (only CFR resulting in the correct answer), children found these tasks to be very difficult at 6 years of age and even later on, as indicated by the current results.

Can executive functioning account for these difficulties? Inhibition (to resist making prepotent but inappropriate responses), cognitive flexibility (to flexibly switch perspectives or strategies), and working memory (to keep information in mind and manipulate it) are the three components of executive functioning being primarily discussed (Diamond, 2006, p. 70).

Inhibitory control is an undeniably important factor for reasoning with assumptions that are counter to fact, as is the case with pretend play (Harris & Kavanaugh, 1993), creating imaginary companions (Taylor, 1999), reasoning with false beliefs (e.g., Wellman, Cross, & Watson, 2001), and future hypothetical reasoning (Robinson & Beck, 2000). Even children who apply BCR need to have sufficient inhibitory strength; otherwise, they would commit the reality bias. Furthermore, inhibitory control predicts logical reasoning abilities in children even when they are older than 4 years (Handley, Capon, Beveridge, Dennis, & Evans, 2004). This can be taken as an indication that successful reasoning not only asks for overcoming the temptation to answer with what is the real state of affairs but also asks for inhibiting "heuristically cued responses" (Handley et al., 2004, p. 192) or individual beliefs that are irreconcilable with the logically valid conclusion.

In the case of CFR, inhibitory control is needed to overcome the temptation of answering with what is the actual state of affairs instead of answering with what would be or would have been the case. A clear sign of children being able to do so is their passing the false belief test. Several studies showed that CFR abilities are related to explicit understanding of false belief (Ecker, 2011; German & Nichols, 2003; Grant, Riggs, & Boucher, 2004; Guajardo & Turley-Ames, 2004; Hofer, 2010; Müller, Miller, Michalczyk, & Karapinka, 2007; Perner, Sprung, & Steinkogler, 2004; Riggs et al., 1998) and that inhibitory control is a partially mediating factor (Drayton, Turley-Ames, & Guajardo, 2011). Inhibition has also been found to predict performance on counterfactual tasks (Beck, Riggs, & Gorniak, 2009), but not whether they experience counterfactual emotions such as regret (Burns et al., 2012).

Although the gain of inhibitory control might serve as an explanation for why children stop committing the reality bias on false belief and counterfactual questions (see also Rafetseder & Perner, 2010), it hardly explains why even older children give wrong answers on our critical conditions. The interference of reality with what needs to be counterfactually assumed is the same among all conditions; the girl comes instead of the boy, and so the candy would be somewhere else than where it really is. So, inhibition of the real events interfering in CFR should be the same, and yet there are some conditions that are much more difficult than others. Children find the condition in which the candy was on the top shelf and they were asked where the candy would be if the little girl had come searching instead of the boy to be much more difficult than the condition in which the candy was on the bottom shelf (and everything else was the same).

Cognitive flexibility, the second component of executive functions, is important for switching between the counterfactual and the real world. It has been reported that counterfactual thoughts increase cognitive flexibility. For example, participants who are primed with a counterfactual mind-set (scenarios that produce counterfactual thoughts) are more likely to solve the Duncker candle task—a measure of flexibility in problem solving—compared with people who were given noncounterfactual primes (Galinsky & Moskowitz, 2000). Cognitive flexibility partially mediates the relationship between performance on counterfactual tasks and that on false belief tasks (Guajardo et al., 2009) and predicts whether children experience regret or not (Burns et al., 2012).

Burns et al. (2012) found that the better 4- to 7-year-olds performed on a task measuring switching abilities, the more likely they were to show counterfactual emotions such as regret. This was interpreted as showing that one needs to switch back and forth between the counterfactual world and the real world to experience regret (Burns et al., 2012, p. 503). Moreover, it is generally agreed that CFR requires two models: one about the actual sequence of events and one about the assumed (hypothesized) sequence of events (Byrne, 2005). People who are using CFR are constantly looking for repercussions of what is possible in the imagined model based on what must be true in the actual model (Perner & Rafetseder, 2011). BCR, in contrast, needs only one model, either about actual events (we have knowledge about the real event and add some new information by applying the laws that hold in this world) or about imagined events (we imagine some event and add new information by applying the laws in this world, e.g., by pretending to wipe up spilled pretend tea with a towel (Harris & Kavanaugh, 1993)). Children who apply BCR might be the ones who actually do not have enough cognitive flexibility to switch between the two worlds and so perform poorly on the critical conditions.

An experiment by Beck et al. (2006) goes against this hypothesis. The 5- and 6-year-old children put out two mats before a mouse came down on one side of a slide that split into two branches halfway down (inverse Y-shape). Thus, they acknowledged that different events can possibly happen at a specific time. Moreover, all children answered correctly that after the mouse had come down on one side (e.g., the left one), the mouse would be on the right side if it had gone the other way. They seemed to grasp that the word "other" needs to be construed as an alternative to a reference point in the actual world (other than the left side), which shows their ability to refer to the real world and construe a possible world in relation to the real world. In short, lack of cognitive flexibility is an unlikely explanation for children's difficulty in applying CFR.

The third component of executive functioning, working memory, has been found to account for the development of conditional reasoning (Barrouillet & Lecas, 1999), a close relative of CFR. With an increasing working memory span, children are able to think of more models in accordance with the conditional interpretation. Santamaría, Espino, and Byrne (2005) supplied evidence that people keep two models ("p and q" as well as "not p and not q") in mind to understand counterfactual statements, whereas they keep only one model ("p and q") in mind to understand indicative statements. It is in fact reported repeatedly that reasoning with indicative statements develops earlier than reasoning with counterfactual statements (Beck et al., 2006; Perner et al., 2004; Rafetseder et al., 2010; Riggs & Peterson, 2000; Riggs et al., 1998; Robinson & Beck, 2000, Study 1).

In line with this, it has been found that the ability to acknowledge multiple possibilities is not fully developed before 6 years of age (Beck et al., 2006). Only at 6 years do most children put out two mats, one at each of the two ends of an inverse Y-shaped slide, in order to make sure that the mouse that comes down the slide does not get hurt no matter which branch it happens to take. Younger children, in contrast, placed out only one mat at the end of one of the branches.

Some studies reported working memory capacity to be related to the amount of counterfactual questions answered correctly (Drayton et al., 2011; Guajardo et al., 2009), but other studies did not confirm this for either counterfactual questions (Beck et al., 2009) or experience of regret (Burns et al., 2012). These studies used different types of counterfactual tasks, which may explain the inconsistency of results. In short, there is no clear-cut relationship between working memory capacity and CFR.

The working memory failure nevertheless might provide a convincing explanation as to why children fail the tasks with discriminating conditions but perform well on the undiscriminating tasks. To be successful, children need to change all of the features of the real world that are causally dependent on the counterfactual assumption but leave everything else the same. Rafetseder and colleagues (2010) referred to this as the nearest possible world constraint. We can think of two explanations as to why children fail to follow this constraint:

(a) *Working memory overload:* Children actually understand that they need to follow this constraint, but their limited working memory capacity prevents them from doing so.
(b) *Lack of understanding the constraint:* Children do not understand that they need to change everything that is causally dependent on the counterfactual assumption but leave untouched everything that is causally independent of that assumption.

Explanation (a) suggests that children fail the task with the discriminating conditions because they need to think of additional information, such as that the girl cannot reach the top shelf or that Max left his shoes on, which is information they do not need to consider with the undiscriminating conditions. This might put a greater burden on their working memory, causing an overload, which results in falling back on BCR and consequently a wrong answer to the test question. Drayton and colleagues (2011) reported that 5-year-old children who performed better on counterfactual tasks had higher working memory scores. In their study, children received antecedent and consequent counterfactual tasks. In the consequent version, they were asked counterfactual questions such as "If there had not been a fire, where would Peter be?" These are tasks that even very young children can solve correctly (Riggs et al., 1998), most likely because they are using simpler reasoning strategies (Rafetseder & Perner, 2010). In the antecedent tasks, children were asked to imagine that they had made the floor dirty. Then, they needed to think of what they could have done to prevent the floor from being dirtied (e.g., take shoes off, wipe off shoes). Children's answers were judged as correct whenever they provided one plausible statement. Again, it is highly likely that children used simpler reasoning strategies (rather than CFR) in order to get this task right. Asking children what one *could have done* in order to prevent the floor from getting dirtied would result in the same answer as when asking children what one *could do* in order to prevent the floor from getting dirty. Thus, from children's answers, we cannot be sure whether they indeed used CFR or whether they applied simple causal reasoning that when somebody takes his or her shoes off (or wipes off the shoes), floors stay clean. What Drayton and colleagues' (2011) results may show is that children who have a higher working memory capacity are the ones who can think of more possibilities, which increases the likelihood of success at least on the antecedent tasks.

Barrouillet and Lecas (1999) similarly found that children's ability to produce more mental models is connected with higher working memory capacity. Participants (8–14 years of age) were told that a child is unsure about how he should get dressed. His mother says, "If you put on a white shirt, then you must put on green trousers." It was the participants' task to find all of the ways the child can get dressed while obeying his mother's rule. Children who produced more cases that were logically consistent with the conditional statement had higher scores on a working memory task (counting span). It was concluded that a higher working memory span enables children to produce a larger number of models underlying a conditional statement.

Taken together, working memory has been shown to be related to the number of models (or possible worlds) one can keep in mind. In both of our tasks, the undiscriminating and discriminating ones, children need to bear in mind two world models: one that represents the real world and another one that represents the counterfactual world. Because the number of worlds is the same for both conditions, this should not account for the differences that we found. However, in the discriminating task, there are extra pieces of information that need to be taken into account (e.g., that the girl is not tall enough in Study 1, that Max walked over the clean floor with his dirty shoes in Study 2). This might place more demands on working memory than is the case in the undiscriminating condition. Memory limitation could explain why children do not keep the fact that the candy was on the top shelf (in Example 1) or on the bottom shelf (in Example 2) in mind when thinking about the counterfactual problem. This would force them to make liberal assumptions from where the protagonist would be fetching the candy (probably the one that comes to their mind first). Although the most obvious first association would still lead to the same and correct answer in Example 2, it would mislead children who assume that the candy is where the girl can reach it (on the bottom shelf) in Example 1.

The current results do not allow us to eliminate this possibility. German and Nichols (2003), however, presented some evidence that this explanation might be true only for 3-year-olds but not for 4-year-olds (and, therefore, older children). They presented children with a chain of events. For example, Mrs. Rosy has planted a new flower and calls her husband to come and look. When Mr. Rosy leaves the house, the dog escapes. The dog runs around the garden and suddenly jumps on the flower, which makes Mrs. Rosy very sad. Children were then either asked to make longer chain inferences (e.g., "What if Mrs. Rosy hadn't called her husband, would Mrs. Rosy be happy or sad?") or a short chain inference ("What if the dog hadn't squashed the flower, would Mrs. Rosy be happy or sad?"). German and Nichols (2003) stated, "For each event in the chain, the outcome of mutating that event results in a proposition that differs from the sequence of events that actually happened" (p. 520). For the longer inference chains, this means that there are more representations of the stored event sequence that need to be considered. In addition, they need to be adapted (i.e., the husband who came out of the house needs to be changed into the husband who stayed in the house and the dog who escaped needs to be changed into the dog who did not escape). The number of representations and their adaptations made a difference for the 3-year-olds who performed above chance on the short chain inference but not on the longer chain inferences. The 4-year-olds, in contrast, performed above chance on all counterfactual questions. We take these results as an indication that children at around 4 years of age have enough working memory capacity to remember and make inferences from a sequence of adapted events. Although Explanation (a) above does seem quite unlikely to explain the differences we found, future studies should include measures for working memory to get a clearer picture of the impact of cognitive demands.

Explanation (b) above suggests that it is not a capacity limitation that keeps children from obeying the nearest possible world constraint. It could be argued that Beck and colleagues' (2006) findings show that 6-year-olds understand this constraint: They answered correctly that a mouse that came down on the left side of a slide would be on the right side if it had gone the other way. So, children changed everything that was causally dependent on the counterfactual assumption (i.e., that the mouse goes the other way) while also referring to the real-world event (i.e., to make sense of the "other" way, one needs to know that this refers to the left side, where the mouse had actually come down). However, we suspect that the word "other" prompted children to search for a reference point in the real world, causing them to fill in missing gaps with specific instances of that world. The real-world event comes into the picture only because of interpreting the word "other" and not as a principle of keeping counterfactual events as close as possible to the actual event and so understanding the constraint. This is yet another indication that children's executive control is good enough to allow them to follow the nearest possible world constraint; however, they lack the understanding that they need to do so.

Alternatively, children in our study might have undone more features of the real world in that possible world than is necessary given by the stated counterfactual. Children might simply have reasoned that if the sister had not played her drums, also grandma would not have dropped the pot. If children undid more facts than they were supposed to undo in Study 2, then they should have done so in Study 1 too. In Study 1, children could have reasoned that if the little girl had come instead of the tall boy, the candy would not have been on the top shelf but instead would have been on the bottom shelf. Reasoning of this kind, however, would suggest that children merely undo facts without paying attention to the causal direction. Although it is reasonable to change an event at Time 2 (the grandma not dropping the pot) based on a changed event at Time 1 (the girl not playing her drums), it is hard to see why one would want to change an event at Time 1 (the candy being on the bottom shelf instead of the top shelf) because an event has changed at Time 2 (the girl coming instead of the boy). See also Lewis (1979) for further discussion on this argument.

How could we distinguish between children who reason on the basis of BCR and children who change too many facts? We would need to change the content of the story such that if Susi had not acted in a certain way there would have been Outcome 1, if Max had not acted in a certain way there would have been Outcome 2, and if neither Susi nor Max had acted in their certain ways there would have been Outcome 3. If children erroneously undid both Susi's and Max's actions when asked a counterfactual question concerning Susi, they should answer with Outcome 3.[2]

In conclusion, our data suggest that the ability to reason counterfactually develops gradually from 6 to 12 years of age. This development is surprisingly late because at around 6 years of age most children are already equipped with a variety of necessary abilities such as keeping two models in mind and switching between these models. This makes it unlikely that children's problems can be explained solely on the basis of weak executive functioning such as working memory, inhibitory control, and/or cognitive flexibility. Future studies, however, should include executive functioning measures (especially for working memory) to be able to draw a definite conclusion. In addition, further studies are needed to find out whether children undo more facts than they are supposed to undo. If future results show that children undo too many facts, then this would still show that they have an incomplete grasp of the nearest possible world criterion. Thus, we conclude that children's problem with counterfactual tasks lies in their lack of understanding of the following principle: One should make only necessary changes to the real world and leave everything else the same (nearest possible world constraint). In particular, one should change only facts that depend logically or causally on the counterfactual assumption. This principle is definitely understood between 6 and 12 years of age.

Acknowledgments

This research was financially supported by an Austrian Science Fund Project (I140-G15), ''Counterfactual reasoning in children,'' as part of the ESF EUROCORES LogiCCC initiative. We are very grateful to Louisa Hacking for her assistance with data collection and to the principals, directors, and children of the following institutions for their cooperation and valuable time in participating in this project: Betriebskindergarten St. Johanns–Spital, Blumenauer Grundschule München, Jugendzentrum IGLU, Jugendzentrum Taxham, Kindergarten Reitberg, NMS Praxisschule der Pädagogischen Hochschule Salzburg, and Volksschule Aigen. We thank the anonymous reviewers for providing us with helpful comments as well as Sarah Zehentner for making corrections.

References

Amsel, E., Robbins, M., Tumarkin, T., Janit, A., Foulkes, S., & Smalley, J. D. (2003). *The card not chosen: The development of judgments of regret in self and others.* Unpublished manuscript, Weber State University.

Amsel, E., & Smalley, J. D. (2000). Beyond really and truly: Children's counterfactual thinking about pretend and possible worlds. In P. Mitchell & K. J. Riggs (Eds.), *Children's reasoning and the mind* (pp. 121–147). Hove, UK: Psychology Press.

Barrouillet, P., & Lecas, J. F. (1999). Mental models in conditional reasoning and working memory. *Thinking and Reasoning, 5*, 289–302.

Beck, S. R., & Crilly, M. (2009). Is understanding regret dependent on developments in counterfactual thinking? *British Journal of Developmental Psychology, 27*, 505–510.

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

Beck, S. R., Robinson, E. J., Carroll, D. J., & Apperly, I. A. (2006). Children's thinking about counterfactuals and future hypotheticals as possibilities. *Child Development, 77*, 413–426.

Burns, P., Riggs, K. J., & Beck, S. R. (2012). Executive control and the experience of regret. *Journal of Experimental Child Psychology, 111*, 501–515.

[2] We thank the anonymous reviewer who provided this suggestion.

Byrne, R. M. J. (2005). *The rational imagination: How people create alternatives to reality*. Cambridge, MA: MIT Press.

Carlson, S. M. (2005). Developmentally sensitive measures of executive function in preschool children. *Developmental Neuropsychology, 28*, 595–616.

Diamond, A. (2006). The early development of executive functions. In E. Bialystock & F. I. M. Craik (Eds.), *Lifespan cognition: Mechanisms of change* (pp. 70–95). Oxford, UK: Oxford University Press.

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.

Ecker, S. (2011). *Kontrafaktisches Denken und Falscher Glaube unter Berücksichtigung der Reasoning-Strategien.* Unpublished master's thesis, University of Salzburg.

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, UK: Oxford University Press.

Epstude, K., & Roese, N. J. (2008). The functional theory of counterfactual thinking. *Personality and Social Psychology Review, 12*, 168–192.

Ferrell, J., Guttentag, R. E., & Gredlein, J. M. (2009). Children's understanding of counterfactual emotions: Age differences, individual differences, and the effects of counterfactual-information salience. *British Journal of Developmental Psychology, 27*, 569–585.

Galinsky, A. D., & Moskowitz, G. B. (2000). Counterfactuals as behavioral primes: Priming the simulation heuristic and consideration of alternatives. *Journal of Experimental Social Psychology, 36*, 384–409.

German, T. P., & Nichols, S. (2003). Children's counterfactual inferences about long and short causal chains. *Developmental Science, 6*, 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*, 177–188.

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

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

Guttentag, R., & Ferrell, J. (2004). Reality compared with its alternatives: Age differences in judgments of regret and relief. *Developmental Psychology, 40*, 764–775.

Guttentag, R., & Ferrell, J. (2008). Children's understanding of anticipatory regret and disappointment. *Cognition & Emotion, 22*, 815–832.

Handley, S. J., Capon, A., Beveridge, M., Dennis, I., & Evans, J. S. B. T. (2004). Working memory, inhibitory control, and the development of children's reasoning. *Thinking and Reasoning, 10*, 175–195.

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

Harris, P. L., & Kavanaugh, R. D. (1993). Young children's understanding of pretense. *Monographs of the Society for Research in Child Development, 58*(1, Serial No. 231).

Hofer, C. (2010). *Kontrafaktisches Denken und Falscher Glaube. Zusammenhänge in der Entwicklung?* Unpublished master's thesis, University of Salzburg.

Kuczaj, S. A., & Daly, M. J. (1979). The development of hypothetical reference in the speech of young children. *Journal of Child Language, 6*, 563–579.

Lewis, D. (1973). *Counterfactuals*. Oxford, UK: Basil Blackwell.

Lewis, D. (1979). Counterfactual dependence and time's arrow. *Noûs, 13*, 455–476.

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.

Obermayr, S. (2011). *Kontrafaktisches Denken unter Berücksichtigung hypothetischer Annahmen.* Unpublished master's thesis, University of Salzburg. .

O'Connor, E., McCormack, T., & Feeney, A. (2012). The development of regret. *Journal of Experimental Child Psychology, 111*, 120–127.

Pearl, J. (2011). The algorithmization of counterfactuals. *Annals of Mathematics and Artificial Intelligence, 61*, 29–39.

Perner, J., & Rafetseder, E. (2011). Counterfactual and other forms of conditional reasoning: Children lost in the nearest possible world. In C. Hoerl, T. McCormack, & S. R. Beck (Eds.), *Understanding counterfactuals, understanding causation: Issues in philosophy and psychology* (pp. 90–109). Oxford, UK: Oxford University Press.

Perner, J., Sprung, M., & Steinkogler, B. (2004). Counterfactual conditionals and false belief: A developmental dissociation. *Cognitive Development, 19*, 179–202.

Rafetseder, E., Cristi-Vargas, R., & Perner, J. (2010). Counterfactual reasoning: Developing a sense of "nearest possible world". *Child Development, 81*, 376–389.

Rafetseder, E., & Perner, J. (2010). Is reasoning from counterfactual antecedents evidence for counterfactual reasoning? *Thinking and Reasoning, 16*, 131–155.

Rafetseder, E., & Perner, J. (2012). When the alternative would have been better: Counterfactual reasoning and the emergence of regret. *Cognition & Emotion, 26*, 800–819.

Riggs, K. J., & Peterson, D. M. (2000). Counterfactual thinking in pre-school children: Mental state and causal inferences. In P. Mitchell & K. J. Riggs (Eds.), *Children's reasoning and the mind* (pp. 87–99). Hove, UK: Psychology Press.

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.

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.

Santamaría, C., Espino, O., & Byrne, R. M. J. (2005). Counterfactual and semifactual conditionals prime alternative possibilities. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 1149–1154.

Skolnick Weisberg, D., & Goodstein, J. (2009). What belongs in a fictional world? *Journal of Cognition and Culture, 9*, 69–78.

Taylor, M. (1999). *Imaginary companions and the children who create them*. Oxford, UK: Oxford University Press.

Weisberg, D. P., & Beck, S. R. (2010). Children's thinking about their own and others' regret and relief. *Journal of Experimental Child Psychology, 106*, 184–191.

Weisberg, D. P., & Beck, S. R. (2012). The development of children's regret and relief. *Cognition & Emotion, 26*, 820–835.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development, 72*, 655–684.

Woodward, J. (2011). Psychological studies of causal and counterfactual reasoning. In C. Hoerl, T. McCormack, & S. R. Beck (Eds.), *Understanding counterfactuals, understanding causation. Issues in philosophy and psychology* (pp. 16–53). Oxford, UK: Oxford University Press.