Journal of Experimental Child Psychology 108 (2011) 190–202

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

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

Supporting children's counterfactual thinking with alternative modes of responding

Sarah R. Beck [a,*], Daniel J. Carroll [b], Victoria E.A. Brunsdon [a,1], Charlotte K. Gryg [a,1]

[a] School of Psychology, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK
[b] Department of Psychology, University of Sheffield, Western Bank, Sheffield S10 2TN, UK

ARTICLE INFO

Article history:
Received 10 May 2010
Revised 16 July 2010

Keywords:
Counterfactual
Executive function
Inhibition
Imagination
Reasoning
Prepotent responses

ABSTRACT

To speculate about counterfactual worlds, children need to ignore what they know to be true about the real world. Prior studies yielding individual differences data suggested that counterfactual thinking may be related to overcoming prepotent responses. In two experiments, we manipulated how 3- to 5-year-olds responded to counterfactual conditional and syllogism tasks. In Experiment 1 ($N = 39$), children's performance improved on both conditional and syllogism tasks when they responded with an arrow rather than pointing with a finger. In Experiment 2 ($N = 42$), 3- and 4-year-olds benefited from both an arrow manipulation and, separately, the introduction of a delay before responding. We suggest that both manipulations help children to overcome an impulsive prepotent response to counterfactual questions arising from a default assumption that information about the past is true.
© 2010 Elsevier Inc. All rights reserved.

Introduction

We often put aside what we know to be true about the world to speculate about what might have been. For example, when a person misses a train, that person imagines the things that he or she could have done to get to the station on time. This is called counterfactual thinking and is pervasive in adult life (Roese & Summerville, 2005). Counterfactual thinking is useful in that it allows us to learn from our mistakes and avoid negative events in the future (Epstude & Roese, 2008). Young children

* Corresponding author.
E-mail address: s.r.beck@bham.ac.uk (S.R. Beck).
[1] The order of these authors was determined alphabetically.

0022-0965/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jecp.2010.07.009

entertain fictional worlds, including pretend ones, from a very young age (Kavanaugh & Harris, 1999; Leslie, 1987). However, it is not until around 4 years of age that they start thinking counterfactually, that is, generating thoughts about what might have been in the real world (e.g., Kuczaj & Daly, 1979; Riggs, Peterson, Robinson, & Mitchell, 1998). This ability continues to develop over several years after this (Beck, Robinson, Carroll, & Apperly, 2006; Rafetseder, Cristi-Vargas, & Perner, 2010; Weisberg & Beck, 2010). In the current article, we consider the relationship between counterfactual thinking and executive function, in particular the question of whether executive demands make counterfactual thinking difficult for preschoolers.

One reason to think that there are executive demands in counterfactual tasks is that children seem to have difficulty in resisting reporting the world as it is rather than as it might be. For example, in Riggs and colleagues' (1998) unexpected transfer task, children were told a story about a person (''Mum'') making a cake. She took some chocolate from the drawer and moved it to the cupboard. When asked, ''What if Mum hadn't made a cake, where would the chocolate be?" 5-year-olds gave the correct counterfactual answer that it would be in the drawer. However, younger children gave a *realist* answer: the cupboard. Thus, they reported the state of the world as it is. Similar shifts from realist answers to counterfactual ones were reported by Harris, German, and Mills (1996) and Guajardo and Turley-Ames (2004), among others.

Riggs and colleagues' (1998) task is a counterfactual conditional task in which the test question takes the form ''If X, then Y?" Evidence from a different type of counterfactual task also suggests that children have difficulty in ignoring what they know about the world. Hawkins, Pea, Glick, and Scribner (1984) found that 4- and 5-year-olds were good at deductive reasoning when premises were congruent with their world knowledge or were about a fantasy world. However, children's performance was very poor when premises were incongruent with their world knowledge (e.g., the premise "birds have wheels"). Children tended to report the truth about the world (''No, birds have wings!"). This tendency to give answers that are true about reality, not the counterfactual world, was also reported by Dias and Harris (1988, 1990) and Richards and Sanderson (1999).

It has been suggested that part of children's difficulty in ignoring reality to give a counterfactual answer may be the result of immature executive function, in particular inhibitory control (Riggs & Beck, 2007; Robinson & Beck, 2000). Inhibitory control is a multifaceted construct, and there has been much discussion over the precise nature of inhibitory control and the situations in which it is deployed (see, e.g., Nigg, 2000). Most explanations involving inhibitory control talk about the need for children to overcome an incorrect but prepotent response. Typical measures of inhibitory control involve situations where participants must select between two possible response options. One option is the correct response, whereas the other response is incorrect but is a stronger candidate for selection—because it is a habitual or well-practiced response, because it is a more desirable response to make, or because it is currently more salient. The stronger but incorrect (i.e., prepotent) response must be inhibited so that the less strongly activated response can be made in its place. Children with immature inhibitory control are more likely to make a prepotent incorrect response, whereas those with stronger inhibitory control can succeed on a task because they are able to suppress the tendency to select the wrong response.

There is a consensus that substantial improvements in inhibitory control occur over the preschool years (e.g., Davidson, Amso, Anderson, & Diamond, 2006), concurrent with improvements in children's performance on counterfactual tasks. Relationships between inhibitory control and counterfactual reasoning have been reported in older children (e.g., Simoneau & Markovits, 2003), and evidence from patients and brain imaging of healthy adults suggests that the orbitofrontal cortex, which is implicated in inhibitory control, is critical for counterfactual thinking (Camille et al., 2004; Coricelli et al., 2005; Ursu & Carter, 2005).

A recent individual differences study found evidence to support this intuition. Beck, Riggs, and Gorniak (2009) tested 3- and 4-year-olds on both counterfactual conditional tasks and syllogism tasks as well as on a battery of executive function measures. Performance on the counterfactual tasks was predicted by inhibitory control, as measured by the Bear/Dragon task (Kochanska, Murray, Jacques, Koenig, & Vandegeest, 1996) and a Stroop-like task (Simpson & Riggs, 2005b). Counterfactual performance was also predicted by language ability but not by working memory. Useful as these data are, this kind of individual differences study (using correlational and regression analyses) can only be suggestive of underlying relationships, and claims about causality cannot be made. Although Beck and colleagues (2009) argued that it was most plausible that developments in domain-general inhibitory control underpinned development of the higher cognitive process of counterfactual thinking, it is of course possible that the reverse is true (see Perner, Lang, & Kloo, 2002, for a similar argument about theory of mind and executive function). A further possibility is that changes in both counterfactual thinking and inhibitory control are being independently driven by an unspecified third process. To make a strong claim about the role of inhibitory control in counterfactual thinking, therefore, we need to directly manipulate the nature of the inhibitory demands of the counterfactual task and observe how this affects children's performance.

Robinson and Beck (2000) made several attempts to manipulate the inhibitory demands in counterfactual conditional tasks. They attempted to increase the salience of the (correct) counterfactual alternative by giving children a representation of it (a picture) and, in another experiment, by making the correct response also the most stereotypical answer to the question. In a scenario where "Teddy" found an object in a videocassette box and swapped it for something else, children found it as difficult to say what would have been in the box had Teddy *not* swapped a videocassette for a fork (so the answer to the counterfactual was what one expects to find in the box) as when Teddy had swapped a fork for a videocassette. Robinson and Beck also attempted to reduce the salience of reality by having the event happen in the past, so that the reality to which the counterfactual referred could no longer be seen in front of the children. None of these manipulations was effective in improving children's performance. Thus, we can rule out one way in which inhibitory demands might impact children's counterfactual thinking. Their problem is not simply that reality is so vivid in their minds that they cannot prevent themselves from referring to it. Such a notion was proposed in Mitchell's (1994) reality masking theory (see also Carlson, Moses, & Hix, 1998).

A different strategy to manipulate executive demands may be useful. In tests of preschool strategic reasoning, a change to the response mode has been found to improve performance (Carlson et al., 1998; Hala & Russell, 2001). One explanation offered to explain this is that alternative response modes may help children to overcome an incorrect prepotent response. For example, Carlson and colleagues (1998) used a deceptive pointing task where children needed to deceive another person about which of two boxes contained a ball. In their experiment, 3-year-olds found this extremely difficult when they gave their response by pointing with a finger, managing to deceive on only 23% of trials. Children showed a strong tendency to give a response based on the true current location of the ball (a kind of realist error). However, performance improved to approximately 60% correct when they were given an arrow to point at the box. Carlson and colleagues argued that veridical pointing is something that children will have practiced and been rewarded for (e.g., when reading picture books with an adult), and so through prior experience it has become a prepotent response. Thus, when asked to point deceptively, children need to overcome the prepotent tendency to point to the actual location of the ball. This places overwhelming demands on 3-year-olds' inhibitory control. On the other hand, when children are given an alternative response mode, such as pointing with an arrow or placing a small marker at a location, children have no previous response history associated with these actions; there is no habitual prepotent response for them to inhibit, and so children are better able to deceive.

There are other accounts of why a nonstandard response mode might improve performance on this and other tasks. Hala and Russell (2001) suggested that alternative response modes allow "cognitive distancing" that enables children to separate the task goal from the means of responding and to symbolically encode the rule on which they need to act. Apperly and Carroll (2009) speculated that by preventing children from responding impulsively, children are able to come up with an alternative response strategy. Once formulated, this strategy then allows children to overcome their previous difficulties with prepotent responses (see also Carroll, Apperly, & Riggs, 2007a,b). We leave further discussion of how a response manipulation might have an effect on counterfactual thinking tasks until after we have demonstrated whether such an effect exists.

In two experiments, we manipulated the response mode in children's counterfactual thinking tasks to investigate the mechanism for how such manipulations might affect counterfactual thinking. Our first experiment tested whether children's performance on counterfactual syllogism and counterfactual conditional tasks is affected by altering the means by which children indicate their response. Our second experiment further explored the influence of nonstandard response modes on counterfactual thinking in preschoolers, with the goal of identifying the cognitive mechanism underpinning a response mode effect. In both experiments, we presented 3- to 5-year-olds with counterfactual thinking tasks and contrasted a traditional response mode of finger pointing with a novel arrow-pointing response. If performance were affected by the response manipulations, this would be evidence that children's difficulty with counterfactual reasoning tasks can be reduced by altering the way they give their response.

Experiment 1

In Experiment 1, we used a rotating arrow manipulation in an attempt to improve children's performance in counterfactual tasks. We used both types of counterfactual task from Beck and colleagues' (2009) individual differences study: counterfactual conditionals and counterfactual syllogisms.

Method

Participants

In total, 19 3- and 4-year-olds (mean age = 3 years 10 months [3;10], range = 3;3–4;2, 10 boys and 9 girls) and 20 4- and 5-year-olds (mean age = 4;9, range = 4;4–5;2, 9 boys and 11 girls) were recruited from and tested at a primary school serving a working-class population of the United Kingdom. Ethnicity was distributed as follows: White 69%, Asian 10%, Black 3%, and other or unknown 18%.

Materials

We used a silver cardboard arrow (20 cm) mounted on a white base using a paper fastener. It could be rotated to the left or right of a midline marked on the base. A toy elephant and toy cow were used in the warm-up trials. Small toys were used to illustrate the counterfactual conditional stories: two dolls, two tables, a shelf, a brown block to represent chocolate, a picture, and a tree. In the syllogism trials, we used picture cards of a purple sheep, a white sheep, a fish in a fish bowl, and a fish in a tree.

Procedure

Children were tested individually in a quiet area of the school. They were alternately assigned to the Arrow or Pointing condition.

Warm-up. In the Arrow condition, children were shown how the arrow moved. The elephant and cow were placed on the table, and children were asked, "Can you point the arrow to the elephant?" and "Now can you show me the cow?" In the Pointing condition, children were asked the same questions but were asked to point with a finger (they were not shown the arrow). No children had difficulty with the warm-up phase.

Experimental trials. There followed two conditional and two syllogism trials presented in pairs of the same type. Due to a procedural error, the order of presentation of trials (conditionals or syllogisms first) was confounded with condition (Pointing or Arrow). This is discussed below.

Conditionals. The conditional stories were taken from Beck and colleagues (2009) and were acted out with toys by the experimenter. At the start of each story, an object was clearly visible in one location; during the course of the story, the object moved to another location, where it was also clearly visible. In the Picture story (based on Riggs et al., 1998), "Jenny" was in the garden painting a picture. She left the picture on the table, and the wind blew it up into a tree. The test question was, "If the wind had not blown, would the picture be on the table or in the tree?" The Chocolate story (based on Wimmer & Perner, 1983) followed the same format. Children scored 1 for pointing to the original location (the correct answer to the counterfactual question). The side on which the correct counterfactual response was placed was counterbalanced.

Syllogisms. The Sheep and Fish syllogism trials were also taken from Beck and colleagues (2009). Children were told that the stories might sound a bit funny, but they were to pretend that they were true (see Leevers & Harris, 2000). For each syllogism, children were first asked a factual check question (e.g., ''What color are sheep?''). If children did not give the expected answer, they were told the correct answer. In the Sheep trial, the picture of the white sheep was placed on the table. Children were then told, ''Let's pretend that all sheep are purple,'' and the corresponding picture was put on the table. Children heard the second premise, ''Michael is a sheep,'' and were asked the test question, ''Is Michael white or purple?'' In the Fish trial, the wording used was, ''Where do fish live? Let's pretend all fish live in trees. Tom is a fish. Does Tom live in a tree or in water?'' The position of the real and counterfactual pictures was counterbalanced between children. Children scored 1 for pointing to the counterfactual picture (the purple sheep and the fish in a tree). In previous studies of this kind, Dias and Harris (1988, 1990) did not use pictures in their tasks, whereas Richards and Sanderson (1999) did. Beck (2007) confirmed that performance was not affected by whether pictures were used.

Results and discussion

Scores were summed to give a total out of two for conditionals and syllogisms (see Table 1). Because there were boys and girls in this sample but only girls in Experiment 2, we checked for gender differences; none was found.

We ran a repeated-measures analysis of variance (ANOVA) with age (3- and 4-year-olds or 4- and 5-year-olds) and condition (Pointing or Arrow) as between-participants factors and trial type (conditional or syllogism) as a within-participants factor. There was a main effect of condition, $F(1, 35) = 7.55$, $p = .009$, partial $\eta^2 = .177$. Children were more likely to give the correct counterfactual response when using the arrow ($M = 1.21$, $SD = 0.58$) than when pointing with a finger ($M = 0.78$, $SD = 0.38$). There were no other significant main effects or interactions (highest $F = 1.00$, lowest $p = .324$). Because the order of trials (conditionals first or syllogisms first) was confounded with the condition (Pointing or Arrow), such that children in the Arrow condition all did syllogisms first, it is possible that the main effect was attributable not to response mode but rather to an order effect. This is very unlikely because there was no difference in performance on the conditionals or syllogisms. Indeed, in Experiment 2 we confirmed that the response mode manipulation had an effect even when it was not confounded with trial type.

To check whether the observed pattern of children's performance differed from that to be expected if children were guessing, chi-square goodness-of-fit tests were conducted. The two age groups were combined to increase power. The results suggested that children were not guessing on the conditional pointing trials, $\chi^2(2, N = 20) = 6.80$, $p = .033$, or on the conditional arrow trials, $\chi^2(2, N = 19) = 10.58$, $p = .005$, and there were borderline differences with chance performance on syllogism pointing trials, $\chi^2(2, N = 20) = 5.90$, $p = .052$, and syllogism arrow trials, $\chi^2(2, N = 19) = 5.95$, $p = .051$. (Note that some expected values for tests of the Arrow condition were 4.75 [i.e., <5]. This level has been deemed as acceptable by some statisticians [e.g., Yarnold, 1970], although we treat these results with caution.)

Children did not give the expected canonical answers (''sheep are white'' and ''fish live in water'') to 29 (of 78) factual check questions for the syllogisms, although the majority of these answers (18) were plausible and conflicted with the counterfactual alternative (e.g., ''sheep are brown,'' ''fish live in a

Table 1
Performance on Pointing and Arrow conditions in Experiment 1.

			Frequency of correct responses			Mean score and standard deviation
			0	1	2	
Younger ($n = 19$)	Conditionals	Pointing	5	3	2	0.70 (0.82)
		Arrow	2	2	5	1.33 (0.87)
	Syllogisms	Pointing	5	4	1	0.60 (0.70)
		Arrow	2	4	3	1.11 (0.78)
Older ($n = 20$)	Conditionals	Pointing	5	3	2	0.70 (0.82)
		Arrow	4	1	5	1.10 (0.99)
	Syllogisms	Pointing	4	1	5	1.10 (0.99)
		Arrow	3	1	6	1.30 (0.95)

Note. Standard deviations are in parentheses.

pond”). We repeated the analysis excluding the seven children who gave implausible answers, and there was no change to the pattern of results.

The use of an alternative response mode appeared to be effective in improving children’s performance on these counterfactual tasks. If we assume that the alternative response mode conveys some kind of executive benefit—perhaps by removing the opportunity for children to make a habitual response—this result is in line with Beck and colleagues’ (2009) individual differences data. However, we wanted to specify more precisely how this mechanism might work. With this in mind, we ran a second study in which we used this manipulation alongside another strategy for improving performance on tasks where children are faced with incorrect prepotent responses.

Experiment 2

Although it was informative to demonstrate that a nonstandard response mode manipulation improved performance on counterfactual thinking tasks, our next challenge was to understand why this should be. The findings from Experiment 1 are consistent with the idea that the arrow helps children to inhibit a habitual response tendency, just as Carlson and colleagues (1998) argued it improved performance on the deception task. However, using an arrow to indicate a box is inherently slower than pointing with a finger, and so another way that the arrow might have an effect is that it might reduce a tendency for children to make an impulsive response, allowing them to make a reflective response to the task and not merely to output the response that happens to be currently prepotent (see Apperly & Carroll, 2009). Our aim in Experiment 2 was to decide between these two different accounts.

The two competing accounts offer different explanations for the locus of the facilitating effect. On the habitual response account, the effect is specific to the means of responding. Any task that uses a response mode that children have habitually used will introduce inhibitory demands when children must use that response in a novel (in this case nonveridical) way. Equally, any task that uses an unusual, novel, or otherwise nonstandard means of responding will not pose these inhibitory demands. In contrast, an account based on impulsivity, such as Apperly and Carroll’s (2009) suggestion, does not see the locus of the effect as being specific to the response mode. Instead, any manipulations that interpose a delay before children are allowed to respond should bring about a facilitating effect regardless of whether children then respond with a habitual response.

Evidence that introducing a delay might improve performance on tasks that make inhibitory demands comes from the Day/Night Stroop task (Diamond, Kirkham, & Amso, 2002). In this task, children see either a sun or a moon picture and are told to say “night” to the sun picture and “day” to the moon picture. Preschoolers find this task to be difficult, with 4-year-olds in Diamond and colleagues’ sample giving the correct response on only 53% of trials. However, when a delay was introduced before children could respond (after each card was turned over, the experimenter sang the ditty, “Think about the answer; don’t tell me!”), performance improved to 88%. Diamond and colleagues speculated that the extra time in their Ditty condition gave the children time to work out a difficult answer on the Day/Night Stroop. An alternative explanation for the same effect would be that obliging children to respond more slowly may prevent children from making an impulsive unreflective response.

A further possible complication when considering the benefit conferred by a delay is that Diamond and colleagues’ (2002) Ditty condition might have encouraged children to focus on the specific rules of the task. Children were explicitly told to think about their answer in the Ditty condition; it may be that the improved performance is attributable to greater focus on the task rules rather than to the delay per se. This interpretation is unlikely given that children’s performance in the original study did not improve when the ditty was sung between trials rather than between question and answer. Nevertheless, Diamond and colleagues noted that the possibility could not be ruled out entirely. Because we wished to look at the specific effect of including a delay, in Experiment 2 we used a procedure entirely separate from the counterfactual task itself to remove this possible confound.

Our main question in Experiment 2 was whether we see comparable improvements in performance on counterfactual tasks when children respond after a delay as when they respond with a rotating arrow. Instead of using a ditty, children were shown a doll and a toy slide and needed to wait until the doll reached the bottom of the slide before responding. Because we found no difference in Experiment 1 between conditionals and syllogisms, we used only counterfactual conditionals in this experiment.

Method

Participants

In total, 20 3- and 4-year-olds (mean age = 3;9, range = 3;4–4;3) and 22 4- and 5-year-olds (mean age = 4;9, range = 4;4–5;3) were recruited from and tested at a primary school serving a middle-class population of the United Kingdom. Ethnicity was distributed as follows: White 64%, Asian 21%, Black 10%, and other or unknown 5%. The school was single sex, and all participants in this sample were girls. The school was fee-paying and selective, meaning that the children were likely to be academically advanced for their chronological age.

Materials

Small toys were used to illustrate the stories: a cat, two trees, a mouse, four dolls, a ball, a mat, a bed, a bench, a shop front, flowers, three tables, a picture, shelves, a brown block (chocolate), a piece of blue paper (pond), and a piece of green paper (garden). For the Delay condition, we used a toy slide ($\sim$30 cm long) and a doll. For the Arrow condition, we used a similar arrow to the one used in Experiment 1.

Procedure

Children were tested individually in a quiet area of the school. Each child participated in all three conditions (Pointing, Arrow, and Delay), and we counterbalanced the order in which the conditions were presented. There were two trials in each condition, and these trials were presented together in pairs.

Warm-up. Warm-ups were presented immediately before each condition. For the Pointing condition, the experimenter placed a pen on the table and asked children to point to it with a finger. For the Arrow condition, the experimenter showed children the movement of the arrow and asked them to point to the pen with the arrow. For the Delay condition, children were told that before they answered the question, they should wait for a doll to come down the slide. Children were shown the doll sliding down the slide, after which they pointed to the pen. The delay before children could respond was approximately 4 s.

Experimental trials. We used six location change stories (see Appendix). Two were used in Experiment 1 (the Chocolate and Picture stories). The other four were devised for the current experiment. In all stories, there were two distinct locations, and a target object moved between these locations during the story. The target object was clearly visible to children at all times. The experimenter described the story to children and acted out the location change with the toys. For example, in the Cat story,

Table 2
Performance on Pointing, Arrow, and Delay conditions in Experiment 2.

		Frequency of correct responses			Mean score and standard deviation
		0	1	2	
Younger (n = 20)	Pointing	8	5	7	0.95 (0.89)
	Arrow	1	3	16	1.75 (0.55)
	Delay	3	4	13	1.50 (0.76)
Older (n = 22)	Pointing	3	2	17	1.64 (0.73)
	Arrow	1	7	14	1.59 (0.59)
	Delay	4	4	14	1.45 (0.80)

Note. Standard deviations are in parentheses.

children were told that the cat was sitting in the tree when it saw a mouse by the pond. The experimenter moved the cat from the tree to chase the mouse at the pond. Children were asked the test question, "If the cat had not spied the mouse, where would the cat be?" Children responded according to the three different conditions; they used a finger to point to the correct location, pointed with the arrow, or needed to delay their answer until the doll had come down the slide and then pointed with a finger. Children scored 1 for indicating the correct counterfactual location (the original location).

Results and discussion

Scores were summed to give a total out of 2 for each condition (Pointing, Arrow, and Delay) (see Table 2).

We ran a repeated-measures ANOVA, making a Greenhouse–Geisser correction because the data violated Mauchley's test of sphericity ($p = .046$), with age (3- and 4-year-olds or 4- and 5-year-olds) and order (there were six possible orders of the conditions) as between-participants factors and condition (Pointing, Arrow, or Delay) as a within-participants factor. There was a significant main effect of condition, $F(1.64, 49.32) = 5.87$, $p = .008$, partial $\eta^2 = .16$, and a significant interaction between condition and age, $F(1.64, 49.32) = 9.02$, $p = .001$, partial $\eta^2 = .23$. There was no main effect of age ($F = 0.91$, $p = .349$). There was no main effect or interaction with order (highest $F = 1.39$, lowest $p = .139$). These results add support to our claim that the main effect in Experiment 1 was most likely due to the difference between conditions (replicated here) and not to the order of presentation.

To investigate the interaction between condition and age, we used post hoc *t* tests to compare performance on each condition across the two age groups, making a Bonferroni correction for three tests ($\alpha = .017$). There was a significant difference between age groups in the Pointing condition, $t(40) = 2.75$, $p = .009$, $r^2 = .16$. The two age groups did not differ in the Arrow ($p = .373$) and Delay ($p = .852$) conditions.

We then ran a separate ANOVA for each age group with condition as a within-participants factor and order as a between-participants factor. For the younger group, there was a main effect of condition, $F(2, 28) = 12.09$, $p < .001$, partial $\eta^2 = .46$. Post hoc *t* tests making a Bonferroni correction for three tests ($\alpha = .017$) showed that children performed worse on the Pointing condition than on both the Arrow condition $t(19) = -4.00$, $p = .001$, $r^2 = .46$, and the Delay condition, $t(19) = -3.24$, $p = .004$, $r^2 = .36$. Arrow and Delay did not differ from each other ($p = .135$). There was no main effect and no interaction with order (highest $F = 1.30$, lowest $p = .276$). For the older group, there were no significant main effects and no interaction (highest $F = 1.33$, lowest $p = .287$). To check whether children's performance was different from what would be expected if they were guessing their answers, chi-square goodness-of-fit tests were conducted. Performances by the older group on all three conditions and by the younger group on the Arrow and Delay conditions were significantly different from chance (all *p*s < .001). Performance by the younger group on the Pointing condition was not significantly different from chance, $\chi^2(2, N = 20) = 5.10$, $p = .078$, so we cannot rule out the possibility that children were guessing. However, we consider this to be unlikely given that 75% of children gave the same response (either both correct or both incorrect) for both trials, suggesting that they were responding systematically and not randomly. We suggest that the lack of difference from chance in this group is due to a lack of power.

The younger children (3- and 4-year-olds) performed poorly when they gave their answers by pointing with a finger. The older children's performance was substantially better than that of both older and younger children in Experiment 1, a result that we believe is most likely due to differences between samples. Children in this sample were recruited from a selective, fee-paying school, and one would expect them to be academically advanced compared with the sample in Experiment 1. Indeed, the older children in Experiment 2 answered correctly on 82% of the "baseline" pointing trials compared with the older children in the Pointing condition in Experiment 1 who answered 35% of the pointing trials correctly. This substantial difference in performance between children of the same chronological age should lead us to be cautious when comparing across samples reported in different articles (e.g., very good performance in Harris et al. (1996) vs. performance in Riggs et al. (1998)). For the current article, the most important point is that the younger children found it easier to give the correct counterfactual answer both when they used the arrow (replicating Experiment 1) and when there was a delay before they responded.

It is worth noting that the delay manipulation was effective even though it was not specifically related to the task at hand. Diamond and colleagues' (2002) ditty directed children, "Think about the answer; don't tell me," and it may have been this reminder to follow the rules of the game rather than the delay that helped children. In the current experiment, the benefit due to the delay manipulation cannot be explained in this way. Thus, although we do not have a direct comparison of a rule reminder delay and our distractor delay, results from Experiment 2 suggest that it is the extra time before responding, and not the meaning of the words themselves, that facilitates children's performance.

General discussion

Theoretical arguments and individual differences data pointed to the possibility that 3- and 4-year-olds' difficulty with counterfactual thinking resulted from difficulties with some aspect of overcoming incorrect prepotent responses. In two experiments, we manipulated aspects of counterfactual tasks that have been argued to reduce these task demands. Both manipulations resulted in improvements to children's performance. In Experiment 1, both 3- and 4-year-olds and 4- and 5-year-olds benefited from using an arrow to give their responses compared with using a finger to point. This was true of both counterfactual conditional trials and counterfactual syllogism trials. In Experiment 2, the 3- and 4-year-olds' performance improved (compared with finger pointing) when they used an arrow and when there was a short delay between the question and their response. In Experiment 2, the 4- and 5-year-olds did not benefit from our manipulations. The most likely explanation for this is that they were already performing at a high level. The group's good performance may reflect the fact that they were an academically able sample.

Our evidence is consistent with the general claim that the developing ability to avoid prepotent responses facilitates children's counterfactual thinking. When considered in conjunction with individual differences data, there emerges a picture that developments in counterfactual thinking are underpinned by domain-general developments in executive function (see Beck, Riggs, & Burns, in press; Riggs & Beck, 2007; but see Sobel, in press). Guajardo, Parker, and Turley-Ames (2009) asked 3- to 5-year-olds to generate antecedent events that would have resulted in a counterfactual consequent. Performance was predicted by working memory and representational flexibility. Because no measure of inhibitory control was included in the study, we cannot be sure whether working memory and representational flexibility make a contribution over and above that of inhibition. Beck and colleagues (2009), whose finding that inhibitory control predicted counterfactual thinking prompted the current study, found that working memory made no unique contribution after controlling for inhibition and language ability (although they did not measure representational flexibility). In older children (up to 7 years of age), Burns, Riggs, and Beck (2010) found that attentional switching predicted children's experience of the counterfactual-based emotion regret. They argued that this was likely because regret arises from a comparison between the real outcome ("I won one sticker") and a counterfactual outcome ("I could have won four stickers") and that such a comparison requires the ability to hold both states in mind and to switch freely between them.

It is implicit in our finding that we can improve children's performance on counterfactual tasks that counterfactual thinking is difficult for young children. Performance on the typical pointing version of the counterfactual tasks reported here was poor, with correct responses on 30 to 55% of trials (with the exception of the older children in Experiment 2). Some researchers have claimed that counterfactual thinking is easy for 3-year-olds (e.g., German & Nichols, 2003; Harris et al., 1996) or even 2-year-olds (Harris, 1997; but see Beck & Guthrie, in press). However, more recently the balance of evidence is in favor of preschoolers performing poorly (e.g., Beck, Riggs, & Gorniak, 2010; Müller, Miller, Michalczyk, & Karapinka, 2007) and that developments continue into middle childhood (Beck et al., 2006; Rafetseder et al., 2010).

We took care when designing our tasks that current reality was on view in all trials (using toys in the conditional trials and pictures in the syllogisms). It is possible that the effects we found are specific to situations in which current reality is visible. It is common to find conditional tasks presented with pictures (e.g., German & Nichols, 2003) or props (e.g., Riggs et al., 1998), but counterfactual syllogism tasks are typically asked verbally without accompanying pictures (although, as mentioned above, Beck

(2007) found no evidence that children performed differently on syllogism tasks with and without pictures). Although it would not be possible to test the arrow manipulation without something for children to point to, it would be useful to extend the findings with the Delay condition to syllogism and condition tasks presented verbally.

An important issue arising from the current findings is to explain how the use of an arrow or a delay might play a role in helping children's counterfactual thinking. In the Introduction, we used Robinson and Beck's (2000) evidence to rule out a reality-masking account. An alternative account came from Carlson and colleagues (1998), who explained the arrow manipulation in terms of children's prior experience. According to their account, a well-practiced and previously rewarded response (in this case veridical pointing) is "extraordinarily difficult for [children] to resist" (p. 674) when children must use it in a nonveridical way. If we accept this premise, it does not seem unreasonable to extend this argument to suggest that difficulties with counterfactual thinking tasks might also be partly attributable to the need to inhibit inappropriate response tendencies arising from children's previous response history. Indeed, at first glance, one might think that this could explain the discrepancy in claims about whether counterfactual thinking is easy for 3-year-olds. Children seemed to perform better in tasks where there was a state change (e.g., from happy to sad, from clean to dirty) and so pointing could not be used to respond (German & Nichols, 2003; Harris et al., 1996), compared with tasks that involved a location change and may have encouraged pointing (e.g., Riggs et al., 1998). However, more recent evidence shows that this is not the case. Beck et al. (2010) directly compared questions about emotions with location change questions and found no consistent differences in performance between the two. It cannot simply be that nonveridical pointing causes children's difficulties on counterfactual tasks.

Further evidence that the demands of executing a task response cannot be the whole story comes from studies where counterfactual trials have been compared with future hypothetical trials and where both tasks require the same kind of response such as Robinson and Beck (2000). In their study, children saw a car travel along a road into a garage at one end. There was a second garage at the other end. Counterfactual questions asked, "What if the car had gone the other way, which garage would it be in?" and were difficult for 3-year-olds, as expected. However, future hypothetical questions, "What if next time he goes the other way, where will he be?" were much easier. Most children correctly pointed at the other garage, not the one where the car actually was. This finding causes problems for both reality masking and habitual response accounts because reality and the prepotent response are the same in both types of trial. Note that Perner, Sprung, and Steinkogler (2004) and Riggs and colleagues (1998) also found that future hypotheticals were easy for 3- and 4-year-olds, but in their tasks reality was different in the counterfactual and future hypothetical versions of the task.

Another reason to think that children's difficulties are not simply with overcoming a prepotent response to point at or report the true state of affairs is that this account would not predict any benefit from the delay manipulation. The fact that a delay is introduced before a response is made should not make any difference as to whether or not a particular response itself is prepotent. Diamond and colleagues (2002) suggested that the delay helped children to pass the Day/Night Stroop task because it gave them time to "compute the answer." This seems plausible within the context of the Day/Night Stroop, but it is a less convincing explanation of why our manipulations helped children with counterfactual tasks. The same computation must be done to reach the correct answer for the future hypotheticals, and this is easy even when there is not a delay.

To explain the facilitating effect of a nonstandard response mode on children's counterfactual thinking, first, it is necessary to set out clearly what we mean by a prepotent response on the counterfactual tasks. Central to this is the suggestion that children's conceptual knowledge *interacts* with their ability to execute a particular response. In this respect, we claim common cause with the account offered by Simpson and Riggs (2005a) to explain the processes involved in children's thinking about false belief. They argued that responses are not inherently prepotent or not prepotent but rather become so as a result of how a task is set up. For example, in the Day/Night Stroop, children find it difficult to say "night" to a picture of the sun and "day" to a picture of the moon, but they find it easy to say "pig" to a picture of the sun and "dog" to a picture of the moon. Thus, a simple prepotent tendency to name the picture on the card cannot be the reason why children find the original condition to be difficult. Instead, Simpson and Riggs argued that context is a crucial factor when considering response prepotency. By their account, responses become prepotent because children plan to say them and set them up as prepared responses ready to be triggered in response to a task-specific stimulus. They speculated that a default assumption in the false belief task—that beliefs are true—could give rise to an incorrect prepotent response. By extension, young children might also have default assumptions about situations featured in counterfactual tasks that create the same sort of prepotent response tendency as on false belief tasks. Specifically, children may have a default tendency to assume that statements about the past (or statements based on general knowledge in the case of syllogisms) are factual (but statements about the future need not be, which would explain the difference between counterfactual and future hypothetical performance). This default assumption about the past may give rise to an incorrect prepotent response in counterfactual tasks.

A common feature of both the Arrow and Delay conditions is that each slows down children's responding. Diamond and colleagues (2002) argued that this time gave children a chance to compute their response on the Day/Night Stroop, but we dismissed this as an explanation for the counterfactual data given the evidence from future hypotheticals. Instead, we think that when asked to respond by pointing with a finger, children are able to do so quickly and unreflectively. This may lead them to make an impulsive response that, because it is unreflective, will be based not on the product of any reasoning about the situation but rather on whichever response is currently prepotent. In contrast, both the delay and arrow manipulations do not allow children to make a rapid response. It is not that they allow children additional time to compute what to do; rather, the enforced pause stops them from outputting a response determined by current prepotency. If this is the case, then we should think of children's executive problems in counterfactual tasks as more a problem of avoiding errors arising from impulsivity than of inhibiting a particular response. One way to add supportive evidence to our claim that the arrow and delay manipulations work in the same way might be to include a combined condition in a future study. If they act on the same problem of impulsivity, then no additional effect of combining them would be expected.

Although the precise role that executive function plays in counterfactual thinking remains to be tested, our experiments have provided direct evidence that this role exists where previously we had only correlational data permitting speculation about such a relationship (Beck et al., 2009). We suggest that children's problems arise from a tendency to respond impulsively and, as a direct consequence of this, to output a prepotent but incorrect response that arises from a default assumption that statements about the past must be truthful. Thus, we are not arguing that counterfactual thinking is easy for young children and that their poor performance results from superficial performance errors; rather, we suggest that default assumptions about what can be said about the past or general knowledge interacts with poor executive control. Children can be supported to give correct answers to counterfactual questions, but until their conceptual understanding develops, they cannot be said to be thinking counterfactually.

Acknowledgments

We thank George Thomas for helpful discussions about this research, and we thank the children and teachers at Rakegate Primary School (Wolverhampton, UK) and Howell's School (Llandaff, UK) for their help and participation.

Appendix

Counterfactual conditional stories used in Experiment 2

1. A cat is sitting in a tree. The cat spies a mouse sitting by the pond. The cat goes to the pond to chase the mouse. If the cat had not spied the mouse, where would the cat be?
2. Penny is playing with a ball on an orange mat. Penny decides she is tired, so she stops playing and goes to bed for a nap. If Penny had not got tired, where would she be?
3. Robin is sitting on a bench when it starts to rain. He goes to the café to stay dry. If it had not rained, where would Robin be?

4. Alice is in the garden and decides to pick some flowers. Alice puts the flowers in a vase on the table. If Alice had not picked the flowers, where would they be?
5. Piglet is drawing a picture at a table in the garden. The wind blows the picture into the tree. If the wind had not blown, where would the picture be?
6. Mum is baking a cake. She takes chocolate from the table and puts it on the shelf. If Mum had not baked a cake, where would the chocolate be?

References

Apperly, I. A., & Carroll, D. J. (2009). How do symbols affect 3- to 4-year-olds' executive function? Evidence from a reverse-contingency task. *Developmental Science, 12*, 1070–1082.
Beck, S. R. (2007). *Relating developments in executive function and counterfactual thinking: Full research report (ESRC End of Award Report, RES-000-22-1683)*. Swindon, UK: Economic and Social Research Council.
Beck, S. R., & Guthrie, C. (in press). Almost thinking counterfactually: Children's understanding of close counterfactuals. *Child Development*.
Beck, S. R., Riggs, K. J., & Burns, P. (in press-a). Multiple developments in counterfactual thinking. In C. Hoerl, T. McCormack, & S. R. Beck (Eds.), *Understanding counterfactuals/understanding causation: Issues in philosophy and psychology*. Oxford, UK: Oxford University Press.
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., Riggs, K. J., & Gorniak, S. L. (2010). The effect of causal chain length on counterfactual conditional reasoning. *British Journal of Developmental Psychology, 28*, 505–521.
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.L., Riggs, K.J., & Beck, S.R. (2010). *Executive control and the development of regret*. Unpublished manuscript, UK: University of Birmingham.
Camille, N., Coricelli, G., Sallet, J., Pradat-Diehel, P., Duhamel, J., & Sirigui, A. (2004). The involvement of the orbitofrontal cortex in the experience of regret. *Science, 304*, 1167–1170.
Carlson, S. M., Moses, L. J., & Hix, H. R. (1998). The role of inhibitory processes in young children's difficulties with deception and false belief. *Child Development, 69*, 672–691.
Carroll, D. J., Apperly, I. A., & Riggs, K. J. (2007a). Choosing between two objects reduces 3-year-olds' errors on a reverse-contingency test of executive function. *Journal of Experimental Child Psychology, 98*, 184–192.
Carroll, D. J., Apperly, I. A., & Riggs, K. J. (2007b). The executive demands of strategic reasoning are modified by the way in which children are prompted to think about the task: Evidence from 3- to 4-year-olds. *Cognitive Development, 22*, 142–148.
Coricelli, G., Critchley, H. D., Joffily, M., O'Doherty, J. P., Sirigui, A., & Dolan, R. J. (2005). Regret and its avoidance: A neuroimaging study of choice behavior. *Nature Neuroscience, 8*, 1255–1262.
Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task-switching. *Neuropsychologia, 44*, 2037–2078.
Diamond, A., Kirkham, N., & Amso, D. (2002). Conditions under which young children can hold two rules in mind and inhibit a prepotent response. *Developmental Psychology, 38*, 352–362.
Dias, M. G., & Harris, P. L. (1988). The effect of make-believe play on deductive reasoning. *British Journal of Developmental Psychology, 6*, 207–221.
Dias, M. G., & Harris, P. L. (1990). The influence of the imagination on reasoning by young children. *British Journal of Developmental Psychology, 8*, 305–318.
Epstude, K., & Roese, N. (2008). The functional theory of counterfactual thinking. *Personality and Social Psychology Review, 12*, 168–192.
German, T. P., & Nichols, S. (2003). Children's counterfactual inferences about long and short causal chains. *Developmental Science, 6*, 514–523.
Guajardo, N. R., Parker, J., & Turley-Ames, K. J. (2009). Associations among false belief understanding, counterfactual reasoning, and executive function. *British Journal of Developmental Psychology, 29*, 681–702.
Guajardo, N. R., & Turley-Ames, K. J. (2004). Preschoolers' generation of different types of counterfactual statements and theory of mind understanding. *Cognitive Development, 19*, 53–80.
Hala, S. M., & Russell, J. (2001). Executive control within strategic deception: A window on early cognitive development? *Journal of Experimental Child Psychology, 80*, 112–141.
Harris, P. L. (1997). On realizing what might have happened instead. *Polish Quarterly of Developmental Psychology, 3*, 161–176.
Harris, P. L., German, T. P., & Mills, P. (1996). Children's use of counterfactual thinking in causal reasoning. *Cognition, 61*, 233–259.
Hawkins, J., Pea, R. D., Glick, J., & Scribner, S. (1984). Merds that laugh don't like mushrooms: Evidence for deductive reasoning by preschoolers. *Developmental Psychology, 20*, 584–594.
Kavanaugh, R. D., & Harris, P. L. (1999). Pretense and counterfactual thought in young children. In L. Balter & C. S. Tamis-LeMonda (Eds.), *Child psychology: A handbook of contemporary issues* (pp. 158–176). Philadelphia: Psychology Press.
Kochanska, G., Murray, K., Jacques, T. Y., Koenig, A. L., & Vandegeest, K. A. (1996). Inhibitory control in young children and its role in emerging internalization. *Child Development, 67*, 490–507.
Kuczaj, S. A., II, & Daly, M. J. (1979). The development of hypothetical reference in the speech of young children. *Journal of Child Language, 6*, 563–579.
Leevers, H. J., & Harris, P. L. (2000). Counterfactual syllogistic reasoning in normal 4-year-olds, children with learning disabilities, and children with autism. *Journal of Experimental Child Psychology, 76*, 64–87.

Leslie, A. M. (1987). Pretense and representation: The origins of "theory of mind". *Psychological Review, 94*, 412–426.
Mitchell, P. (1994). Realism and early conception of mind: A synthesis of phylogenetic and ontogenetic issues. In C. Lewis & P. Mitchell (Eds.), *Children's early understanding of mind: Origins and development* (pp. 19–45). Hove, UK: Lawrence Erlbaum.
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.
Nigg, J. T. (2000). On inhibition/disinhibition in developmental psychopathology: Views from cognitive and personality psychology and a working memory taxonomy. *Psychological Bulletin, 126*, 220–246.
Perner, J., Lang, B., & Kloo, D. (2002). Theory of mind and self-control: More than a common problem of inhibition. *Child Development, 73*, 752–767.
Perner, J., Sprung, M., & Steinkogler, B. (2004). Counterfactual conditionals and false belief: A developmental dissociation. *Cognitive Development, 19*, 179–201.
Rafetseder, E., Cristi-Vargas, R., & Perner, J. (2010). Counterfactual reasoning: Developing a sense of "nearest possible world". *Child Development, 81*, 376–389.
Richards, C. A., & Sanderson, J. A. (1999). The role of imagination in facilitating deductive reasoning in 2-, 3-, and 4-year-olds. *Cognition, 72*, B1–B9.
Riggs, K. J., & Beck, S. R. (2007). Thinking developmentally about counterfactual possibilities. *Behavioral and Brain Sciences, 39*(5–6), 463.
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. R. (2000). What is difficult about counterfactual reasoning? In P. Mitchell & K. J. Riggs (Eds.), *Children's reasoning and the mind*. Hove, UK: Psychology Press.
Roese, N., & Summerville, A. (2005). What we regret most ···and why. *Personality and Social Psychology Bulletin, 31*, 1273–1285.
Simoneau, M., & Markovits, H. (2003). Reasoning with premises that are not empirically true: Evidence for the role of inhibition and retrieval. *Developmental Psychology, 39*, 964–975.
Simpson, A., & Riggs, K. J. (2005a). Factors responsible for performance on the Day–Night task: Response set or semantics? *Developmental Science, 8*, 360–371.
Simpson, A., & Riggs, K. J. (2005b). Inhibitory and working memory demands of the Day–Night task in children aged 3–11 years. *British Journal of Developmental Psychology, 23*, 471–486.
Sobel, D.M. (in press). Domain-specific causal knowledge and children's reasoning about possibility. In C. Hoerl, T. McCormack, & S. R. Beck (Eds.), *Understanding counterfactuals/understanding causation: Issues in philosophy and psychology*. Oxford, UK: Oxford University Press.
Ursu, S., & Carter, C. S. (2005). Outcome representations, counterfactual comparisons, and the human orbitofrontal cortex: Implications for neuroimaging studies of decision-making. *Cognitive Brain Research, 23*, 51–60.
Weisberg, D., & Beck, S. R. (2010). Children's thinking about their own and others' regret and relief. *Journal of Experimental Child Psychology, 106*, 184–191.
Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*, 103–128.
Yarnold, J. K. (1970). The minimum expectation for χ^2 goodness-of-fit tests and the accuracy of approximations for the null distribution. *Journal of the American Statistical Association, 65*, 864–886.