CORRECTED PROOF

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

British Journal of Developmental Psychology (2009), in press
© 2009 The British Psychological Society

The British Psychological Society

www.bpsjournals.co.uk

The effect of causal chain length on counterfactual conditional reasoning

Sarah R. Beck[1]*, Kevin J. Riggs[2] and Sarah L. Gorniak[1]
[1]University of Birmingham, Birmingham, UK
[2]London Metropolitan University, London, UK

We investigated German and Nichols' finding that 3-year-olds could answer counterfactual conditional questions about short causal chains of events, but not long. In four experiments ($N = 192$), we compared 3- and 4-year-olds' performance on short and long causal chain questions, manipulating whether the child could draw on general knowledge to answer. We failed to replicate German and Nichols' result, finding instead that in two experiments (Experiments 1 and 3) there was no difference in performance on short and long causal chain questions and in two experiments (Experiments 2 and 4) children showed the opposite pattern: short causal chain questions were more difficult than long. These two unexpected patterns of results were replicated in a fifth study ($N = 97$). Children with lower language ability found short causal chains more difficult than long. Performance by children with higher language ability was unaffected by the length of the causal chain they had to consider. We found no evidence that children showed precocious counterfactual thinking when asked about recent events in a causal chain and conclude that counterfactual thinking develops after 4 years of age.

Counterfactual thinking is the ability to consider alternatives to what we know has happened. Such an ability may be related to thinking about others' minds and reasoning about causes. Understanding *when* counterfactual thinking develops is important for understanding these relationships. Yet, there remains some controversy as to when this development occurs.

German and Nichols (2003) claimed that counterfactual thinking is an early development. In their study, 3-year-olds performed well on their most simple counterfactual task. Children heard stories about a chain of events. For example, Mrs Rosy planted a flower in the garden and called her husband to see. He opened the door from the house, allowing the dog to escape. The dog ran in to the garden and squashed the flower, making Mrs Rosy sad. Three-year-olds answered more than two-thirds of counterfactual questions correctly when they had to change the most recent event in

* *Correspondence should be addressed to Dr Sarah R. Beck, School of Psychology, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK (e-mail: s.r.beck@bham.ac.uk).*

DOI:10.1348/026151009X450836

CORRECTED PROOF

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

the chain and were asked about the emotional consequences for the character, 'What if the dog hadn't squashed the flower, would Mrs Rosy be happy or sad?'. Questions that referred to an event further back in the story, for example, 'What if Mrs Rosy hadn't called her husband, would Mrs Rosy be happy or sad?' were substantially more difficult for 3- and 4-year-olds, with the younger children answering only 15% correctly. The authors argued that success on the short causal chain questions is evidence for 3-year-olds' counterfactual thinking.

Other studies have also reported evidence of counterfactual thinking in very young children. Harris, German, and Mills (1996) showed children short, simple chains of events acted out with dolls. For example, in one Carol came in from the garden wearing muddy boots and walked across the floor, leaving dirty footprints. Children were asked, 'What if Carol had taken her shoes off, would the floor be dirty?'. Sixty-nine per cent of their sample of 3-year-olds answered counterfactual questions about all or all but one of four scenarios correctly. Harris (1997) also showed that children as young as 2 years old appeared to be able to understand the word 'almost' as referring to a counterfactual world. Having seen one toy horse gallop right to the edge of the table and stop abruptly and another stop safely some distance from the edge, they could identify that it was the first that 'almost fell'. Finally, Perner, Sprung, and Steinkogler (2004) found that 3-year-olds were able to answer counterfactual conditional questions about a simple scenario in which each of two antecedents led to only one consequent (e.g. if you take the boat you go to the lake, if you take the coach you go to the pasture). The view that counterfactual thinking is early developing has become widely accepted among developmental psychologists as well as cognitive psychologists interested in theoretical accounts of counterfactual thinking (e.g. Byrne, 2005; Harris, 2000).

There is, however, reason to doubt this widely held view. There is evidence that young children find tasks that have the same structure as short causal chain counterfactuals difficult. For example, Riggs, Peterson, Robinson, and Mitchell (1998; see also Robinson & Beck, 2000) showed children a sorting task. Objects were sorted in to one of two trays based on whether or not they were decorated with a picture. The experimenter drew on a piece of paper and placed it in the pictures tray. 3- and 4-year-olds found it relatively difficult to answer the counterfactual question, 'If I had not drawn on the piece of paper, which box would it be in?'. Guajardo and Turley-Ames (2004) also found that 3- and 4-year-olds found counterfactual thinking difficult, both in tasks where the child was given a counterfactual antecedent and asked to provide the consequent (like those used by Riggs *et al.*, 1998) and when children were given a counterfactual consequent and asked to generate an antecedent. For example, having heard a story modelled on Harris' muddy boots scenario, children were asked, 'What could you have done so the kitchen floor would not have gotten dirty?'.

Children's understanding of almost has recently been questioned. Although children can identify which horse almost fell when the alternative choice came nowhere near falling, in an almost identical condition where the other horse actually fell off the edge of the table 3-year-olds performed poorly. In one experiment, they systematically indicated the horse on the floor when asked, 'Which horse almost fell?' (Beck & Guthrie, 2009).

Thus, there is a serious discrepancy in the literature as to whether 3-year-olds are competent counterfactual reasoners. The claim made by German and Nichols (2003) and Harris *et al.* (1996) that 3-year-olds can speculate about what might have been is important, not only for our understanding of children's hypothetical thinking,

CORRECTED PROOF

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

but also in relation to their developing theory of mind (see also Riggs *et al.*, 1998; Guajardo & Turley-Ames, 2004). The German and Nichols (2003) study is particularly important in this debate because they found a difference between the versions of their task depending on the length of the causal chain of events the child had to consider. Based on this, they argue that even the youngest children were able to do counterfactual thinking. Of course, the relative ease with which 3-year-olds answered German and Nichols' short causal chain questions poses a problem for those that argue that children of this age have great difficulty with counterfactual thinking. However, there is an alternative explanation of German and Nichols' (2003) data that suggests that 3-year-olds' success on the short causal chain questions are false positives.

This alternative explanation (see Beck, Robinson, Carroll, & Apperly, 2006; Perner, 2000) argues that children do not need to use counterfactual thinking to answer the short causal chain questions correctly. Instead it may be possible to answer these questions using one's general knowledge of the world. For example, when flowers are squashed people tend to be sad, when flowers are unsquashed people tend to be happy. Similarly, balloons make people happy, and burst balloons make people sad (see German & Nichols, 2003, Balloon Story). Importantly, this general knowledge explanation could potentially explain other evidence of precocious counterfactual thinking. In Harris *et al.* (1996), Carol walks across a floor with muddy boots, leaving dirty footprints. Children are asked about the state of the floor had Carol taken her shoes off. General knowledge tells us (and perhaps 3-year-olds) that dirty boots lead to a dirty floor, and no boots lead to a clean floor.

Perner's critical test for whether a question could be answered with general knowledge, and hence without counterfactual thinking, is whether the right answer can be reached without hearing the preceding story. Children could give the correct answer, 'happy' to the question, 'What if the dog hadn't squashed the flower, would Mrs Rosy be happy or sad?' without hearing the story. But it is not clear what the answer to the long causal chain question 'What if Mrs Rosy hadn't called her husband, would Mrs Rosy be happy or sad?' is without hearing the specific events of this story. What looks like precocious counterfactual thinking could in fact be the child drawing on her general knowledge.

Thus, we pitched this new explanation against the original precocious counterfactual thinking explanation (i.e. that short causal chains reveal early counterfactual thinking). To do this we devised new stories which permitted short and long causal chain questions that could not be answered using general knowledge. These stories contained a mishap (which may be important to prompt counterfactual thought, see German, 1999) that led to both a change in emotional state for a character and a change in location of a key object. For example, in one story closely modelled on German and Nichols' flower story, a dog stole a spade from the boy playing in the sandpit and dropped it in the pond making the boy sad. Our second story was not so closely based on German and Nichols' second story (the balloon story) because it was difficult to include a change in location. In our new vase story a football knocks a vase off a windowsill which smashes on the floor. We asked four different kinds of counterfactual question (examples here are from the new spade story): short emotion questions (What if the dog hadn't stolen the spade, would Tom be happy or sad?) long emotion questions (What if Tom hadn't called his Dad [who opened the door allowing the dog to escape], would Tom be happy or sad?), short location questions (What if the dog hadn't stolen the spade, would the spade be in the sandpit or in the pond?), and long location questions (What if Tom hadn't called his Dad, would the spade be in the sandpit or in the pond?).

German and Nichols' (2003) precocious counterfactual thinking explanation predicted that children should find both short emotion and short location questions

CORRECTED PROOF

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

easy, as they involve the same degree of inferential processing. Long emotion and long location questions should both be difficult. In contrast, the general knowledge explanation (Beck *et al.*, 2006; Perner, 2000) predicted that only the short emotion questions should be easy, as these were the only questions that could be answered with general knowledge: stealing things makes people unhappy, as does smashing vases, and the converse can be assumed to make people happy. However, there is no *a priori* reason to think that Tom would be happy or sad if he had not called his Dad (long emotion), or to know where the spade would be if Tom had not called his Dad (long location), or it had not been stolen (short location).

EXPERIMENT 1

In Experiment 1, we tested children on our four question types: the short and long emotion questions based on German and Nichols (2003) and our new short location and long location questions. We anticipated finding evidence for either the precocious counterfactual thinking explanation (all short questions easier than all long questions) or for the alternative general knowledge explanation (only short emotion questions easier than all other questions).

Method

Participants

Fifty children (29 girls) mean age 4 years and 4 months (range 3 years and 10 months to 4 years and 9 months) participated in the study. The sample comprised children from a variety of ethnic backgrounds, with the largest group (48%) being Black British, and from a school serving a predominantly working class population. In this and all subsequent studies children either spoke English as a first language, or were deemed capable of understanding the task by the class teacher.

Design

Children were alternately assigned to one of two conditions. Each child was read two stories each with two counterfactual questions. Half the children were asked one short emotion question and one long location question for each story. The other half of the children were asked one short location and one long emotion question for each story. The order of the stories (spade or vase first), the order of the causal chain question (short or long question first), and the order of the forced choice options (would the character be 'happy or sad', or 'sad or happy') were counterbalanced. This produced 16 different combinations.

Materials

We used four pictures (approximately 21 × 15 cm, mounted on laminated card 30 × 22 cm) to illustrate each story. Examples of the pictures we used are shown in Figure 1.

Procedure

Children were tested individually in a corner of the classroom. We used two new stories based closely on the materials used by German and Nichols (2003). In the *spade story* (modelled on German and Nichols' Mrs Rosy story), Tom had built a sandcastle.

CORRECTED PROOF

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

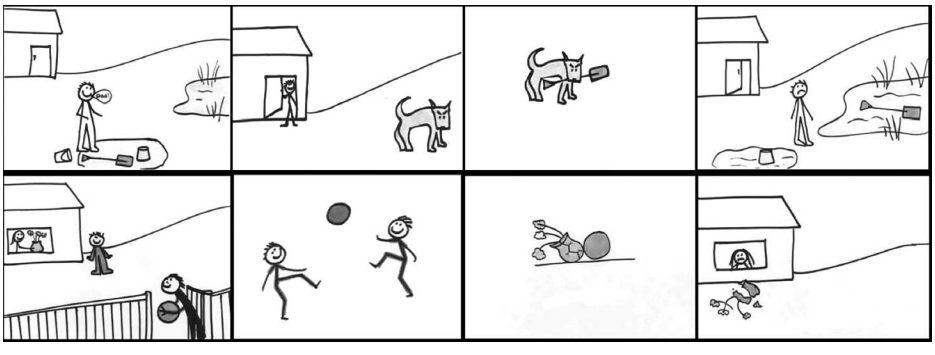

Figure 1. Pictures used to illustrate our new Spade (top row) and Vase (bottom row) stories. In Experiment 1, the third picture in each sequence also included the main character in the picture (Tom in the sandpit and Nicholas kicking the football).

He called his father to see it (picture 1). When his father opened the door the dog escaped from the house (picture 2), ran into the garden, stole Tom's spade (picture 3), and dropped it in the pond. This made Tom sad (picture 4). The four pictures were placed in a row on the table as the story progressed. In the *vase story*, Nicholas' mother puts some flowers in a vase on the windowsill. Nicholas' friend comes to play football (picture 1). Nicholas kicks the ball too hard (picture 2) and it knocks over the vase. The vase breaks (picture 3) and Nicholas' mother is sad (picture 4). Scripts are included in the Appendix. Note that in both stories a mishap occurred, which should encourage counterfactual thinking (see e.g. German, 1999).

After each story, each child was asked six questions in two blocks of three. First there were two control questions as in German and Nichols (2003). Prior to emotion test questions children were asked, 'Just now is Tom happy or sad?' and 'Right at the beginning was Tom happy or sad?'. These were followed by either a short counterfactual test question, 'What if the dog hadn't stolen the spade, would Tom be happy or sad?' or a long counterfactual test question, 'What if Tom hadn't called his Dad, would Tom be happy or sad?'. Prior to the two location test questions the control questions were, 'Just now is the spade in the sandpit or the pond?' and 'Right at the beginning was the spade in the sandpit or the pond?'. The short counterfactual test question in this condition was, 'What if the dog hadn't stolen the spade, would the spade be in the sandpit or in the pond?'. The long counterfactual test question was, 'What if Tom hadn't called his Dad, would the spade be in the sandpit or in the pond?'.

The experimenter pointed to the corresponding picture when asking each question (following German and Nichols' procedure). Children scored 1 for each question they answered correctly. These were summed to give a score out of 8 for the control questions, a score out of 2 for the short counterfactual questions (half the children had emotion questions, half had location), and a score out of 2 for the long counterfactual questions (half the children had location questions, half had emotion)[1].

[1] *We compared children's performance on the two stories for each type of question. There were 16 comparisons over the four experiments reported in this paper. Making a strict Bonferroni correction for 16 tests* $p = .003$*, none of the comparisons were significant. Considering borderline differences, there was one occasion where more children passed the spade story than the vase story (Experiment 2, long location questions* $N = 26$*,* $p = .004$*). On one occasion more children passed the vase story than the spade story (Experiment 4, long emotion questions* $N = 19$*,* $p = .070$*). There were no other significant differences between the stories (lowest p value was .29). We did not pursue these item differences further, believing them to be spurious.*

CORRECTED PROOF

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

Results and discussion

We conducted a repeated measures ANOVA with causal chain length (short, long) as a within subjects factor and condition (Short emotion and long location, short location and long emotion) as a between subject factor. If the precocious counterfactual thinking explanation was correct and short counterfactuals were consistently easier than long counterfactuals (German & Nichols, 2003) then there would be a main effect of causal chain length. On the other hand, if short counterfactuals were unusually easy because children could answer them using general knowledge (the alternative explanation), there would be an interaction between causal chain length and condition. There were neither main effects nor interactions (highest $F = 1.21$, lowest $p = .28$). Children's performance by question type can be seen in Table 1.

Six children made an error on a control question. In this and all subsequent experiments, we repeated the analysis excluding those children who made any errors on control questions. This never changed the pattern of results.

We had found no support for the general knowledge explanation that short emotion causal chain questions were artefactually easy for children to answer as there was no interaction between condition and causal chain length. However, we also found no support for German and Nichols' (2003) precocious counterfactual thinking explanation. Short causal chain questions were not easier than long causal chain questions. This was surprising, as we had assumed that we would replicate the original finding. It was presumptuous to abandon both explanations based on only one sample of children, so we tested a second sample in Experiment 2. Our aim remained to seek evidence for either the general knowledge explanation or the precocious counterfactual thinking explanation.

EXPERIMENT 2

In Experiment 2, we used the same stories and questions as in Experiment 1 with a new sample of children.[2] However, having noted that there were some differences between our pictures and those used by German and Nichols (2003) we made some minor changes to the materials. We realized that in the original German and Nichols' study the pictures were somewhat smaller than our pictures (12 cm × 10 cm). We wondered if children found it difficult to view all four of our large (30 cm × 22 cm) pictures simultaneously when thinking about the story. Thus we reduced the size of our pictures so they were easier to view (17.5 cm × 12.5 cm). We also made a minor change to the wording of our counterfactual test questions saying 'had not' rather than 'hadn't' which brought our questions in line with those used by Riggs *et al.* (1998) and would make the negation of the event clearer for the children (note that in the German & Nichols study children were tested in their native language, Greek, so we could not take the exact wording from their paper).

Method

Participants

Fifty-two children (29 girls) mean age 4 years and 4 months (3 years and 10 months to 4 years and 9 months) participated in the study. The sample comprised children from a variety of ethnic backgrounds, with the majority (94%) being White British, and from a school serving a predominantly working class population.

[2] *No child participated in more than one experiment.*

CORRECTED PROOF

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

Table 1. Frequencies of correct answers to each type of question by experiment

		Short causal chain number correct (only children who passed controls)				Long causal chain number correct (only children who passed controls)			
Experiment	Question type	0	1	2	Mean	0	1	2	Mean
Experiment 1 ($N = 50$)	Emotion	12 (9)	3 (3)	11 (11)	0.96	4 (4)	10 (9)	10 (10)	1.25
	Location	6 (6)	10 (9)	8 (8)	1.08	7 (6)	9 (7)	10 (10)	1.12
Experiment 2 ($N = 52$)	Emotion	6 (2)	4 (3)	16 (16)	1.38*	5 (3)	8 (4)	13 (12)	1.31 (ns, $p = .058$)
	Location	14 (11)	10 (6)	2 (2)	0.54*	3 (2)	9 (7)	14 (12)	1.42**
Experiment 3 ($N = 47$)	Emotion	9 (2)	4 (3)	11 (9)	1.08	8 (2)	10 (6)	5 (3)	0.87
	Location	6 (3)	6 (2)	11 (6)	1.22	9 (3)	5 (4)	10 (7)	1.04
Experiment 4 ($N = 41$)	Emotion	11 (4)	8 (4)	3 (2)	0.64*	4 (1)	8 (1)	7 (3)	1.10
	Location	8 (2)	8 (2)	3 (0)	0.74	6 (1)	10 (4)	6 (5)	1.00
Study 5									
Low BPVS ($N = 47$)		25 (24)	11 (9)	11 (10)	0.70*	10 (10)	18 (16)	19 (17)	1.19
High BPVS ($N = 50$)		16 (16)	9 (9)	25 (25)	1.18	18 (18)	12 (12)	20 (20)	1.04

*$P < .05$; **$P < .01$.

Note. Mean scores are based on all children. Comparisons with chance were made using one sample *t* tests, thus a significant difference indicates that the group mean is different to chance. Non significant findings can represent that individual children were guessing (the pattern apparent in Experiment 1 short location) or that individual children were consistent but roughly equally distributed between passers and failers (the pattern apparent in Experiment 1 short emotion).

CORRECTED PROOF

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

Design

This remained the same as in Experiment 1.

Materials

As in Experiment 1, we used four pictures to illustrate each story. Their size was reduced (approximately 17.5 cm × 12.5, mounted on laminated card 21 cm × 15.5 cm) to enable children easily to view them simultaneously. Their size was now comparable to that in German and Nichols (2003; 10 cm × 12 cm). We made changes to some of the pictures to ensure they were comparable across stories and similar to the original German and Nichols' materials: Picture 3 in both stories was changed to show only the action being mentioned (i.e. the dog stealing the spade and the ball hitting the vase) so that children were not given a clue about the characters' expression or the objects' location in this picture. In Experiment 1, the first picture in the spade story showed the spade in Tom's hand. We changed this so it was resting in the sandpit to ensure that the correct answer to the control and counterfactual location questions was unambiguous.

Procedure

The procedure was identical to that in Experiment 1, except that the wording of the test questions was changed from 'what if . . . hadn't . . . ' to 'what if . . . *had not* . . . ' for clarity.

Results and discussion

We conducted a repeated measures ANOVA with causal chain length (short, long) as a within subjects factor and condition (short emotion and long location, short location and long emotion) as a between subjects factor. We found a main effect of causal chain length $F(1, 50) = 11.29$, $p = .002$, $\eta_p^2 = .18$, a main effect of condition $F(1, 50) = 7.97$, $p = .007$, $\eta_p^2 = .14$, and an interaction between causal chain length and condition $F(1, 50) = 9.24$, $p = .004$, $\eta_p^2 = .16$. We used *post hoc t* tests to investigate this interaction, making a Bonferroni correction for four tests $p = .0125$. There was no difference between conditions on the long causal chain questions. On short causal chain questions performance was better in the short emotion/long location condition ($M = 1.38$) than the short location/long emotion condition ($M = 0.54$), $t(50) = 4.03$, $p < .001$. Furthermore, children in the short emotion/long location condition performed equally well on short and long questions, whereas children in the short location/long emotion question performed better on long ($M = 1.31$) than short ($M = 0.54$) questions $t(25) = -4.81$, $p < .001$.

Our design meant that the same children did not have both long and short questions of the same type. We used independent *t* tests to make these comparisons. Short location questions ($M = 0.54$) were more difficult than long location ($M = 1.42$), $t(50) = 4.72$, $p < .001$, but there was no difference in difficulty between short and long emotion questions.

We repeated the analysis excluding 12 children who made a mistake on a control question. There were no changes to the pattern of significant main effects, interactions nor *post hoc* tests.

In Experiment 2, we did find a difference between performance on questions with varying causal length. However, this was in the opposite direction to that predicted by the precocious counterfactual thinking explanation. Our new short location question was

CORRECTED PROOF

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

substantially more difficult to answer than either short emotion questions or long location questions. This finding offers potential support for the general knowledge explanation.

However, one problem for the general knowledge explanation was that children performed surprisingly well on long causal chain questions. We wondered if both types of long question were open to a false positive response. When children were asked the long question the experimenter pointed to the first picture in the sequence. This clearly depicted the correct 'counterfactual' answer to the test questions: the protagonist happy and the key object in its original location. The corresponding picture for the short question did not show the correct answer. In Experiment 3, we modified the procedure to prevent these potential false positives. Thus, we sought more conclusive evidence for the general knowledge explanation.

EXPERIMENT 3

Our aim in Experiment 3 was to remove the possibility that good performance on long questions represented false positives, resulting from children 'reading off' the answer from the first picture in the sequence. In order to prevent this, we simply ensured that the earlier pictures were not visible when the test questions were asked. Pictures were presented as the story was read to the children, but each was placed on top of the preceding one.

If the general knowledge explanation is correct then short emotion questions should now be easier than all other questions (because these are the only ones that can be answered without reference to the story). Neither Experiment 1 nor 2 provided support for the precocious counterfactual thinking explanation, so we did not make predictions based on it in Experiment 3 (both short questions easier than both long).

Method

Participants

Forty-eight children (24 girls) mean age 4 years and 6 months (3 years and 11 months to 4 years and 10 months) participated in the study. One boy was excluded as he did not appear to understand the task. The sample comprised children from a variety of ethnic backgrounds, with the majority (65%) being Black British, and from a school serving a predominantly working class population.

Design and materials

These remained the same as in the previous experiment.

Procedure

This followed the procedure for Experiment 2 but with the following modification: the pictures were laid down on top of one another when the children were being read the story, rather than in a line across the table. Only the last picture (picture 4) was visible when the control and test questions were asked, and no pictures were pointed to when the questions were asked. This prevented children from giving a correct answer to the long counterfactual questions by describing the picture the experimenter pointed to when answering the question.

CORRECTED PROOF

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

Results and discussion

We conducted a repeated measures ANOVA with causal chain length (short, long) as a within subjects factor and condition (short emotion and long location, short location and long emotion) as a between subjects factor. There were no significant main effects or interactions (highest $F = 1.48$; lowest $p = .23$).

More children made errors on control questions in this experiment (47%) when the pictures were turned over than in either Experiment 1 (8%) or 2 (15%). This is unsurprising as the pictures were not on view to prompt their memory. We repeated the analysis excluding these children and there was no change to the pattern of results.

Experiment 3 provided no further evidence for the general knowledge explanation (nor for precocious counterfactual thinking). However, we also noted that seemingly slight changes in the methodology had a marked effect on children's ability to answer control questions. Thus we thought it prudent to confirm our results using German and Nichols' (2003) original materials. In Experiment 4, children completed a replication of one of German and Nichols' trials and we used our stories once more, but with attempts to reduce the memory demands.

EXPERIMENT 4

We used our own stories with the short and long emotion and location questions to make a final test of the general knowledge and precocious counterfactual thinking explanations. We modified the procedure to reduce the memory demands: pictures were laid out in a row as the story was read and were turned over after a recap. We included an extra trial using German and Nichols' own story and design to confirm that our failure to find evidence for the precocious counterfactual thinking explanation was not due to our materials used in Experiments 1–3.

Method

Participants

Forty-three children (24 girls) mean age 3 years 6 months (3 years and 0 month to 4 years and 1 month) participated in this study. The sample comprised children from a variety of ethnic backgrounds, with the majority (72%) being White British, and from a school serving a predominantly working class population.

Design

This remained the same as in the previous experiments.

Materials

We used the same pictures as in Experiments 2 and 3. We used a third set of four pictures based on German and Nichols' description of their Balloon story. We did not use the Mrs Rosy story because it was too similar to our own adapted spade story.

Procedure

The procedure followed the same format as Experiment 2. However, once each story was finished, the experimenter recapped the events for each picture turning them over as she

CORRECTED PROOF

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

did. Control questions and test questions were asked once all the pictures were turned over.

After children had heard the spade and the vase stories they heard a third story, German and Nichol's Balloon story, in which Nicholas is playing outside with a balloon. His mother calls him inside. He leaves the balloon with his friend, who unfortunately falls into a rosebush, popping the balloon. Nicholas is sad. Children are asked either the short (emotion) causal chain question, 'What if his friend hadn't fallen in the rosebush. Would Nicholas be happy or sad?' or the long (emotion) causal chain question, 'What if Mum hadn't called Nicholas in for snack. Would he be happy or sad?' (We changed mother to mum for our British participants).

Results and discussion

First we analysed the data from our spade and vase stories. We conducted a repeated measures ANOVA with causal chain length (short, long) as a within subjects factor and condition (short emotion and long location, short location and long emotion) as a between subjects factor. There was a main effect of causal chain length $F(1, 39) = 9.52$, $p = .004$, $\eta_p^2 = .20$. Performance was worse on short causal chains ($M = 0.69$) than on long causal chains ($M = 1.08$). There were no other significant effects or interactions (highest $F = 0.44$, lowest $p = .51$). As in Experiment 3 children could not read the correct answer to the long questions off the pictures, yet performance on these questions was better than the short. There was no support for our concern following Experiment 2 that good performance on long questions resulted from false positives, where children simply described the first picture.

A Wilcoxon signed ranks test showed that in the subset who answered all control questions correctly (33%) there remained a difference between short and long questions $Z = -2.50$, $p = .013$. There was no difference between short and long emotion questions (Mann–Whitney $U = 16$, $p = .64$) or short and long location questions (Mann–Whitney $U = 18.5$, $p = .84$). We checked this result including another 10 children who made only one mistake on the control questions (59% of the original sample). The pattern was the same.

Our data on the balloon story, an exact replication of German and Nichols' procedure, was out of line with their findings. Eight children (of 22) passed the short causal chain question whereas 11 (of 19) passed the long causal chain question. This was not significant $\chi^2 = 1.90$, $p = .17$. We excluded children who failed the control questions for this story (leaving $N = 36$ children). The pattern of results was the same: 8 (of 20) children passed the short causal chain and 10 (of 16) passed the long causal chain. Again this was not significant $\chi^2 = 1.80$, $p = .18$.

In Experiment 4, we found no evidence to support either the precocious counterfactual thinking or the general knowledge explanations. We also remained concerned that small changes to methodology appeared to cause substantial problems with children's recall and potentially their counterfactual thinking.

STUDY 5

In Experiments 1–4, we had contrasted two possible explanations for German and Nichols' (2003) finding that counterfactual questions about short causal chains were easier to answer than those about long causal chains. First, there was their own

CORRECTED PROOF

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

explanation that children's success on short questions reflected genuine precocious counterfactual thinking. Second, others (e.g. Beck *et al.*, 2006; Perner, 2000) had suggested that children could answer some questions using general knowledge, not counterfactual thinking. We found no evidence to support either of these hypotheses. Thus we rejected both these explanations.

Our most striking finding is that we never replicated German and Nichols' original finding that short counterfactuals were easier to answer than long. Instead, we found two different and unexpected patterns of results: (i) no difference between short and long questions, (ii) short questions were easier than long. Our experimental manipulations could not account for these different patterns, thus we wondered whether they may have resulted from differences in the samples. In our final study, we used two individual differences measures: age and linguistic ability. Both have been observed to relate to children's counterfactual reasoning ability in previous studies (Guajardo & Turley-Ames, 2004; Riggs *et al.*, 1998). While we could not offer any *a priori* explanation for our two unexpected patterns of results, we hoped that patterns of performance in subsamples defined by age and linguistic ability may shed light on the causes of these patterns. We investigated this possibility using data from another study (Beck, Riggs, & Gorniak, 2009) of a large sample of approximately 100 children. This allowed us to compare older and younger children and more and less linguistically able children within the same sample, which was not possible across the smaller samples in Experiments 1–4. We used German and Nichols' own stories. Having rejected the general knowledge explanation, we had no reason to think that our stories offered any advantage over the original published materials.

Participants

There were 97 children in the sample (48 girls) mean age 3 years and 11 months (3 years and 3 months to 4 years and 5 months (dates of birth were not available for 2 children)). The sample comprised children from a variety of ethnic backgrounds, with the majority (74%) being White British, and from schools serving working and middle class populations.

Materials

We used German and Nichols' Mrs Rosy and Balloon stories. We used 4 pictures (approximately 17.5 cm × 12.5, mounted on laminated card 21 × 15.5 cm) to illustrate each story, based on the descriptions in their paper. We used the British Picture Vocabulary Scale Second edition (BPVS; Dunn, Dunn, Whetton, & Burley, 1997) to assess receptive vocabulary.

Procedure

Children were tested on four counterfactual, four executive function tests and the BPVS over two sessions lasting approximately 25–30 min each. Here we focus on only the data from the causal chains task and the BPVS, as the relationships among the counterfactual measures and between counterfactual and executive measures are reported elsewhere (Beck *et al.*, 2009). Children had first participated in a working memory task (Gordon & Olson, 1998). They then completed the BPVS and immediately after they completed the causal chains measure.

CORRECTED PROOF

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

In the BPVS, the child sees four pictures on each trial. The experimenter reads a target word and the child has to point to the appropriate picture. Testing continues until children make eight mistakes in a block of 12 words.

Our causal chains test used the stories and questions devised by German and Nichols. One difference from the original study was that comparison between causal chain length was within, rather than between subject. As in Experiments 1–4 children were asked both short and long questions after each story. All children heard the Mrs Rosy story followed by the Balloon story (in the original paper the order of the stories was also fixed). After the experimenter read the story and set out the sequence of pictures, children were asked two control questions (e.g. 'Just now is Mrs Rosy happy or sad?' and 'Right at the beginning was Mrs Rosy happy or sad?') and then either the short causal chain question, e.g. 'What if the dog hadn't squashed the flower, would Mrs Rosy be happy or sad?' or the long causal chain question, e.g. 'What if Mrs Rosy hadn't called her husband, would Mrs Rosy be happy or sad?' Children were then asked a filler question, 'What colour is Mrs Rosy's top [blouse]?' before being asked the other causal chain question. The filler question was included so that children would not have two counterfactual questions to which the same answer was correct immediately following each other. All children answered the filler question correctly. Following German and Nichols' procedure, and because we had not found any evidence that children's performance resulted from false positives, pictures were left on display while children were asked the test questions.

Results and discussion

We used median splits to form groups of younger (39–46 months, $M = 43$, $SD = 2$, $N = 46$) and older (47–53, $M = 50$, $SD = 2$, $N = 49$) children and groups of low BPVS score (15–40, $M = 33$, $SD = 5$, $N = 47$) and high BPVS score (41–71, $M = 49$, $SD = 7$, $N = 50$). We checked that there were no differences in performance on the two stories for short ($p = .26$) or long ($p > .99$) questions and combined the scores across the two items. We ran a repeated measures ANOVA with causal chain length (short or long) as a within subjects variable and age (younger or older), BPVS (low or high) and order (short or long question first) as between subject variables.

There were no significant main effects. However, there was a significant interaction between causal chain length and BPVS group $F(1, 87) = 5.05, p = .027, \eta_p^2 = .058$ (next highest $F = 2.82$, $p = .097$, BPVS × order). Making a Bonferroni correction for four comparisons, *post hoc t* tests showed that children with low BPVS scores performed worse on short causal chains ($M = 0.70$) than long causal chains ($M = 1.19$), $t(46) = -3.16, p = .003$ (the pattern seen in Experiments 2 (location questions) and 4). Children who scored highly on BPVS did not discriminate $t(49) = 0.87$, $p = .391$ (the pattern seen in Experiments 1 and 3). Both low and high scoring children performed equally well on long causal chains, $t(95) = 0.90$, $p = .37$, but high scoring children ($M = 1.18$) out-performed low scoring children ($M = 0.70$) on the short causal chains $t(95) = -2.72, p = .008$. Thus, children with relatively low receptive vocabulary found the short causal chain questions particularly difficult.

Four children made errors on the control questions. We repeated our analysis with these children excluded and there was no change to the pattern of significant effects and interactions.

CORRECTED PROOF

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

GENERAL DISCUSSION

We began this research because we were interested in why children found short causal chain questions easier to answer than long causal chain questions (German & Nichols, 2003). Two alternative explanations had been proposed: that children's success on the short causal chain questions revealed their precocious counterfactual thinking (German & Nichols, 2003) or that they were false positives reached by general knowledge, rather than counterfactual reasoning (Beck *et al.*, 2006; Perner, 2000). Our data did not support either of these two explanations.

The reason we could not endorse either of these explanations was that in five studies, with 289 children, we never replicated the original advantage of short over long causal chains (see also Chan & Hahn, 2007). This was the case even when we used German and Nichols' materials (Study 5) and exactly the same design, i.e. one test question per story (Experiment 4, Balloon story). We can offer no explanation for their findings. We note that there are cultural and linguistic differences between their sample (Greek children) and ours (British children), but it is not clear why these differences should lead to the different patterns of results. Instead, we focus on what our data can tell us about children's counterfactual thinking.

First, our data support the claim that counterfactual thinking is relatively difficult for 3- to 4-year-old children. German and Nichols (2003) found ceiling performance by the 4-year-olds they tested on the short causal chains. Our samples never reached this level of success. Our data are in line with other authors who report that counterfactual conditionals are relatively difficult for preschoolers and are understood in their fifth year (e.g. Riggs *et al.*, 1998).

Our data fit a picture of counterfactual thinking developing over early and middle childhood. The first sign that children think about alternative worlds appears before their second birthday, when they begin to pretend. By the fifth birthday children are able to answer conditional questions about specific counterfactual events. Still to develop is the understanding that the alternative world was once a possibility that could have happened instead of what really did happen (at around 5 or 6 years Beck *et al.*, 2006), and the understanding of the emotional consequences of considering these counterfactual worlds, such as regret or relief at around 7 years (Guttentag & Ferrell, 2004; see also Beck & Crilly, 2009). Those who have claimed that children can do counterfactual thinking from 3 years of age have not directly addressed these later developments. One possibility would be to suggest that they result from incidental task demands, not genuine developments in counterfactual thinking. However, based on the evidence we think it is more likely that thinking counterfactually has a relatively late and protracted development.

Second, our data shed light on how individual children approach counterfactual tasks. In our studies, there were two distinct patterns of performance: either there was no difference between short and long causal chain questions, or short questions were more difficult than long. The reanalysis of our individual differences data (Study 5) exposed both these patterns within the same sample of children. Differences in children's linguistic ability predicted which pattern they showed. Children who performed well on our measure of linguistic ability (the BPVS) did not discriminate between the two types of questions. Children who did more poorly on this measure performed worse on the short than on the long questions. A simple interpretation of our data is that children with better language abilities try to engage in counterfactual thinking for both short and long causal chains. They may find this difficult, but they find

CORRECTED PROOF

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

it equally difficult if the counterfactual world imagined is several steps back in the causal chain or just one.

In contrast, the children with poorer language may not have understood the wording of the counterfactual test questions. Thus, rather than try to consider an alternative past world, one strategy they may have adopted was to track back in time to the point referred to and report on the state of affairs then. This strategy would produce the pattern of results we saw: the long causal chain question directs him or her back to the start of the story and the correct answer to the counterfactual question (e.g. Mrs Rosy happy was (not would have been) happy). Note that this is different to the false positive explanation for our data investigated in Experiments 3 and 4 that children could simply read off the picture. Instead, children may be remembering the beginning of the story. However, remembering a past event is not counterfactual thinking. The short causal chain question directs the child to the most recent event when there was no explicit mention of how the character felt at that point. This is no help in answering the short question.

Why might language ability predict children's ability to handle counterfactual tasks? It may be (as we have hinted above) that the BPVS measure is acting as a proxy for general cognitive development, and this general ability enables children to engage with the task. It might also be that counterfactual questions make substantial linguistic demands on children. Perhaps, these are specific in the same way that false belief reasoning has been related to children's mastery of particular grammatical structures (see e.g. De Villiers & Pyers, 2002; Smith, Apperly, & White, 2003) or they may be general. It is difficult to think of a counterfactual task that does not require language to cue one to imagine what might have been, and it might be that general language development underpins children's ability to create complex models of the world.

In conclusion, our data suggest that once children are able to engage with counterfactual questions their performance is not dependent on the length of the chain of events they have to consider. We should be cautious about accepting claims that children develop counterfactual thinking very early. The results of our experiments are more in line with the view that developments in counterfactual thinking continue after 4 years.

Acknowledgements

This research was funded by a grant (RES000221683) from the Economic and Social Research Council to the first and second authors.

We thank the children and teachers at Lillian de Lissa Nursery, Shenley Fields Nursery, St Thomas' Centre, Weoly Castle Nursery, King David Nursery, King's Norton Nursery, Castle Vale Nursery, and Marsh Hill Nursery Schools, who participated in this study.

References

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., & Guthrie, C. (2009). Almost thinking counterfactually: Children's understanding of close counterfactuals. Manuscript submitted for publication.

Beck, S. R., Riggs, K. J., & Gorniak, S. L. (2009). Relating developments in children's counterfactual thinking and executive functions. Manuscript submitted for publication.

CORRECTED PROOF

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

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*(2), 413–426.

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

Chan, A. & Hahn, B. (2007). Causal order effect in three- and four-year-olds' counterfactual reasoning. Poster Presented to the Biennial Meeting of the Society for Research in Child Development, Boston, MA: USA.

De Villiers, J. G., & Pyers, J. (2002). Complements to cognition: A longitudinal study of the relationship between complex syntax and false-belief understanding. *Cognitive Development*, *17*, 1037–1060.

Dunn, L. M., Dunn, L. M., Whetton, C., & Burley, J. (1997). *British picture vocabulary scale* (2nd ed.). London: NFER Nelson.

German, T. P. (1999). Children's causal reasoning: counterfactual thinking occurs for 'negative' outcomes only. *Developmental Science*, *2*, 442–447.

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

Gordon, A. C. L., & Olson, D. R. (1998). The relation between acquisition of a theory of mind and the capacity to hold in mind. *Journal of Experimental Child Psychology*, *68*, 70–83.

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.

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

Harris, P. L. (1997). On realizing what might have happened instead. *Polish Quarterly of Developmental Psychology*, *3*, 161–176.

Harris, P. L. (2000). *The work of the imagination*. Oxford: Blackwell.

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

Perner, J. (2000). About + Belief + Counterfactual. In P. Mitchell & K. J. Riggs (Eds.), *Children's reasoning and the mind*. Hove: Psychology Press.

Perner, J., Sprung, M., & Steinkogler, B. (2004). Counterfactual conditionals and false belief: A developmental dissociation. *Cognitive Development*, *19*, 179–201.

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: Psychology Press.

Smith, M., Apperly, I. A., & White, V. (2003). False belief reasoning and the acquisition of relative clause sentences. *Child Development*, *74*(6), 1709–1719.

Received 22 May 2008; revised version received 30 March 2009

Appendix

Script

Spade story

Picture 1: 'Here is Tom Rosy. He's just built a sandcastle and he's very happy with it. He calls his Dad, Mr Rosy from the house to come and have a look.'

CORRECTED PROOF

Copyright © The British Psychological Society
Reproduction in any form (including the internet) is prohibited without prior permission from the Society

Picture 2: 'When Mr Rosy opens the door to come into the garden, the dog escapes from the kitchen.'

Picture 3: 'The dog runs around the garden. Look he steals Tom's spade and drops it in the pond!'

Picture 4: 'Now the spade is stuck in the middle of the pond, and Tom is sad.'

Test questions:

Short emotion: 'What if the dog had not stolen the spade, would Tom be happy or sad?'

Short location: 'What if the dog had not stolen the spade, would the spade be in the sandpit or in the pond?'

Long emotion: 'What if Tom had not called his Dad, would Tom be happy or sad?'

Long location: 'What if Tom had not called his Dad, would the spade be in the sandpit or in the pond?'

Vase story

Picture 1: 'Nicholas' mother is putting some flowers in a vase. She puts the vase on the windowsill. She thinks the flowers are very pretty. She is very happy. Nicholas' friend has come to play with a ball.'

Picture 2: 'Nicholas and his friend play football in the garden.'

Picture 3: 'Nicholas kicks the ball too hard. Look! It knocks over the vase.'

Picture 4: 'Now the vase is broken and the flowers are on the floor. Nicholas' Mother is sad.'

Short emotion: 'What if the ball had not broken the vase, would Nicholas' Mother be happy or sad?'

Short Location: 'What if the ball had not broken the vase, would the flowers be in the window or on the floor?'

Long Emotion: 'What if Nicholas' friend had not brought the ball, would Nicholas' mother be happy or sad?'

Long Location: 'What if Nicholas' friend had not brought the ball, would the flowers be in the window or on the floor?'

(In Experiment 1, question was 'hadn't', rather than 'had not'. This was changed to improve clarity.)