Thinking & Reasoning
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/ptar20

Relating developments in children's counterfactual thinking and executive functions

Sarah R. Beck [a], Kevin J. Riggs [b] & Sarah L. Gorniak [c]

[a] School of Psychology, University of Birmingham, UK

[b] Department of Psychology, London Metropolitan University, UK

[c] School of Psychology, University of Birmingham, UK

Available online: 26 Oct 2009

To cite this article: Sarah R. Beck, Kevin J. Riggs & Sarah L. Gorniak (2009): Relating developments in children's counterfactual thinking and executive functions, Thinking & Reasoning, 15:4, 337-354

To link to this article: http://dx.doi.org/10.1080/13546780903135904

THINKING & REASONING, 2009, 15 (4), 337–354

Psychology Press
Taylor & Francis Group

Relating developments in children's counterfactual thinking and executive functions

Sarah R. Beck
School of Psychology, University of Birmingham, UK

Kevin J. Riggs
Department of Psychology, London Metropolitan University, UK

Sarah L. Gorniak
School of Psychology, University of Birmingham, UK

The performance of 93 children aged 3 and 4 years on a battery of different counterfactual tasks was assessed. Three measures: short causal chains, location change counterfactual conditionals, and false syllogisms—but not a fourth, long causal chains—were correlated, even after controlling for age and receptive vocabulary. Children's performance on our counterfactual thinking measure was predicted by receptive vocabulary ability and inhibitory control. The role that domain general executive functions may play in 3- to 4-year olds' counterfactual thinking development is discussed.

Keywords: Counterfactual thinking; Imagination; Inhibition; Reasoning; Working memory.

When we miss trains or have to work late to meet a deadline, thoughts of how the world might have been otherwise come to us. In the last decade developmental psychologists have explored what occurs when young children come to imagine alternatives to events that have happened. They have used different tasks with variation in the task structures and presentation. A more systematic approach is called for if we are to

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

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

We thank Elizabeth Robinson and Jackie Blisset for their advice on this paper, and the children and teachers at King David Nursery, King's Norton Nursery, Castle Vale Nursery, and Marsh Hill Nursery Schools, UK, who participated in this study.

http://www.psypress.com/tar DOI: 10.1080/13546780903135904

understand how children come to adult-like counterfactual thinking. Thus our first aim in this paper is to test whether these tasks share a common counterfactual element. Our second aim is to investigate whether domain-general developments in executive functions underpin children's ability to think counterfactually.

In counterfactual conditional tasks children hear about a series of events and are asked to imagine how the world would be now, given a false antecedent. For example, in a task used by Riggs, Peterson, Robinson, and Mitchell (1998) Mum was making a cake. She took the chocolate from the drawer, used some, and put it away in the cupboard. Children were asked, "What if Mum hadn't made a cake, where would the chocolate be?" Until children were about 4 years of age they answered these questions incorrectly, reporting the current location rather than where the chocolate was earlier. These counterfactual conditional tasks vary in whether the event involves a mishap and how the tasks are presented (with props or pictures), and there is some discussion about the relative difficulty of these tasks (German & Nicholls, 2003; Guajardo & Turley-Ames, 2004; Harris, German, & Mills, 1996; Perner, Sprung, & Steinkogler, 2004).

There is another type of counterfactual reasoning task. Counterfactual syllogisms involve the participant reasoning with a premise that they know to be false to derive a valid conclusion. For example, "All cats bark. Penny is a cat. Does Penny bark?" Although the empirically true answer is that cats (and so Penny) do not bark, if one treats the first premise as if it were true then the logically valid answer is "yes". Children tested on these tasks (e.g., Dias & Harris, 1998, 1990; Leevers & Harris, 2000) found it remarkably difficult to reason about these false premises, and they tended to give the empirically correct answer until they were around 6 years old. However, when they were encouraged to use their imaginative powers and mental imagery, performance improved. Children gave the logically correct counterfactual answers from about 4 years of age (if not earlier, see Richards & Sanderson, 1999).

To date there has been no study that has included both types of counterfactual tasks, conditionals and syllogisms, which might be expected to rely on a shared ability. Indeed, there have not been comparisons between different counterfactual conditional tasks. Authors draw their conclusions based on only one measure (e.g., German & Nichols, 2003; Riggs et al., 1998). Thus by including a variety of counterfactual reasoning tasks in our battery we provide the first evidence as to whether they do or do not share a common element.

We now turn to our second aim: What general processes might be related to explicit counterfactual thinking? For several reasons it has been suggested that executive functions are likely to play a role in counterfactual thinking. These "control" processes undergo substantial development at around the same time as children move from making realist errors on the counterfactual tasks described above (e.g., Espy, McDiarmid, Cwik, Stalets, Hamby, & Senn, 2004). Furthermore, executive processes have been linked to later developments in reasoning (e.g., Handley, Capon, Beveridge, Dennis, & Evans, 2004; Simoneau & Markovits, 2003). Although there is limited neuroscience research in this area, there is some evidence (Ursu & Carter, 2005) and theoretical reasons (Baird & Fugelsang, 2004) to think that executive functions and counterfactual thinking recruit similar areas of the prefrontal cortex. In a related area of cognitive development, there are many reports that theory of mind performance is related to executive functions (e.g., Carlson and Moses, 2001; Flynn, 2007; Perner, Lang, & Kloo, 2002) and there are several authors who report strong relations between theory of mind and counterfactual thinking (e.g., Perner et al., 2004; Riggs et al., 1998). Finally, authors studying counterfactual thinking itself have noted that executive functions such as working memory or inhibitory control are exactly the sorts of developments that *could* account for improvement in counterfactual thinking (e.g., Byrne, 2005; Robinson & Beck, 2000).

Within the realm of executive function, both working memory and inhibitory control are plausible candidates for facilitating counterfactual thought. In the case of working memory it is possible that imagining an alternative, counterfactual state of affairs requires one to generate new information about that counterfactual world, while still holding in mind information about the real world. With regard to inhibitory control, when answering questions about the counterfactual world one has to resist responding with what one knows to be true. The need to ignore how the world is, in favour of how it might be, is likely to make demands on children's inhibitory control.

As we mentioned above, we are not the first to investigate the relationship between domain general executive functions and reasoning. There is good evidence that inhibition and working memory play important roles in logical reasoning for both adults (e.g., Markovits & Doyon, 2004), children in late childhood (8 years and older), and adolescents (de Neys & Everaerts, 2008; Handley et al., 2004; Simoneau & Markovits, 2003). Some of the tasks require participants to generate their own conclusions. In others their task is a recognition one: participants are provided with a possible conclusion and must evaluate it. In some of the studies participants are asked to reason with premises known to be false and to judge what (if anything) can be inferred from them. Why, then, is it of interest to investigate the relationship between counterfactual thinking and executive functioning in younger children?

One reason is that the children studied by Riggs, Harris, German, and colleagues on the simple counterfactual tasks are substantially younger than any described in the studies above. We do not know whether the same relationships between counterfactuals and executive measures exist at such a young age when counterfactual thinking is first emerging. Second, the tasks that have received the most attention from researchers interested in the emergence of counterfactual thinking are somewhat different from those used in the studies above. In Riggs et al.'s (1998) location change task, and German and Nichols' (2003) causal chains task, children hear a narrative describing a specific event. Counterfactual reasoning is used to change one of the components of the narrative and to infer a probable outcome. On the other hand, the false syllogism task (Dias & Harris, 1988) asked preschoolers to deduce a conclusion from a set of preferences, one of which was known to be false. In the adolescent and adult literatures, work on counterfactual thinking about events (e.g., Byrne, 2005) and logical counterfactual reasoning (e.g., Handley et al., 2004) have also been treated separately, and tend to be of interest to different groups of researchers. Using both types of task will allow us to see whether the narrative conditional tasks are related to, and make the same general demands as, the syllogism task for very young children.

Thus, in our study we compared children's performance on various counterfactual thinking tasks with performance on two types of executive function task: inhibitory control and working memory. We tested 3- and 4-year-old children on our battery of tasks to find out whether different counterfactual measures tap the same construct and if children's ability to reason counterfactually is related to measures of executive function.

METHOD

Participants

A total of 107 children (56 boys) from nurseries serving working- and middle-class populations participated in the study—mean age 3 years and 10 months (3;10), $SD = 3.9$, range 3;2–4;5. Of these, 14 children were not included in our final analyses as they did not complete the second session of testing (due to refusal or absence from school), leaving 93 children in the final sample (45 boys, mean age 3;11, $SD = 3.9$, range 3;3–4;5). Two further children refused to complete the syllogisms task, which was presented last, and dates of birth were not available for two different children. Data from these participants were included in comparisons where possible. All children spoke English as a first language, or were deemed competent in English by their class teacher. A total of 74% of the children were Caucasian, 12% were Black, 11% were Asian, 3% were of other or mixed ethnic origin.

Procedure

Children were tested individually in two 25–30-minute sessions, with 2–10 days between sessions. The tasks were presented in a fixed order, as is typical in individual differences research, as follows: Session 1 = Counting & Labelling (Working Memory), British Picture Vocabulary Scale (second edition), Causal chain counterfactuals, Bear Dragon (Inhibitory Control); Session 2 = Noisy Book (Working Memory), Location change conditionals, Black/ White Stroop (Inhibitory Control), Counterfactual syllogisms. Using a fixed order is standard practice in individual differences research. It allows fair comparison of the data from different children as they experience the tasks in the same context as each other. However, it limits comparisons between tasks. Without a counterbalanced design one cannot say for sure that any difference between two tasks is due to intrinsic differences between them or other artefactual influences, such as fatigue (see Carlson & Moses, 2001, for further discussion). As our goal was to look for relationships between measures using an individual differences approach and we had no a priori reason to think that any particular order would influence children's performance, we used a single fixed order of presentation.

Measures

Verbal ability measure

The British Picture Vocabulary Scale Second edition (BPVS; Dunn, Dunn, Whetton, & Burley, 1997) is a standardised test measuring receptive vocabulary. The child pointed to one of four pictures that s/he judged best fit the target word spoken by the experimenter.

Inhibitory control measures

Bear/dragon. This is a child-friendly go/no go task (Kochanska, Murray, Jacques, Koenig, & Vandegeest, 1996). Children imitated 10 actions that the experimenter demonstrated (e.g., "touch your nose"). The child was told that when "nice bear" talked they were to do what he said (go trials), but when "naughty dragon" told them to do something they should not do it (no go trials). There were two practice trials with feedback, one with each puppet. If the child responded incorrectly the experimenter demonstrated the correct response. The experimenter restated the rules before the test phase.

In the test phase there were 16 instructions presented in a pseudo-random order—Bear (B), Dragon (D), BBDDBDBDBDDBBD. After eight trials the rules were restated. Performance on the go trials was excellent. Only three children made mistakes on these trials (one child made two mistakes, two children made one mistake). Our measure was based on performance on the inhibitory no go trials. Children scored 1 each time they did not respond to the dragon (maximum = 8).

Black/white Stroop. In this test (Gerstadt, Hong, & Diamond, 1994; Simpson & Riggs, 2005) children were shown a white card and told "When you see this card, I want you to say black". Equivalent instructions were given for the black card ("say white"). After the child responded the rule was recapped, "Remember, when you see this card, you don't say black, you say white", and vice versa.

In the test phase there were 16 cards presented in a pseudo-random order—Black (B), White (W), BWWBBWBWWBWBBW. If children did not respond to a card they were prompted, "What do you say for this card?" Children scored 1 each time they said the other colour name (maximum = 16).

Working memory measures

Counting and labelling. In this test (Gordon & Olson, 1998) three items (car, sock, pencil) were placed before the child, and three items (crayon, necklace, bear) before the experimenter. The experimenter named then counted her items and asked the child to do the same: "I've got a crayon, necklace, bear. Can you name your toys? ... Now I'm going to count them. One, two, three. Can you count your toys?" The experimenter then said, "In this game we are going do both together like this: One is a crayon, two is a necklace, three is a bear. Can you do the same with your toys?" If children failed on the first attempt the experimenter repeated the counting and labelling combined with her items, and asked the child to try again. In a second test trial the child's original set of objects was replaced by a new set (bike, cup, hat). The full procedure was repeated.

On each trial children scored 2 if they successfully counted and labelled the toys on the first attempt and 1 if they succeeded on the second attempt. Scores were summed (maximum = 4).

Noisy book. In this test (Hughes, 1998) children were presented with a 3 × 3 grid of small pictures, taken from a Red Riding Hood story book (Nolan, 1993), mounted on an electronic board. Each picture made a different noise when pressed. Children named and pressed each picture to familiarise them with the noises and the locations. The experimenter tested recognition of the pictures by asking the child to press each of them in turn. The pictures were then covered with a piece of paper with a 3 × 3 grid drawn on it. In a practice trial children were asked to press two pictures (horn, frogs) in that order. Children were given feedback on this practice attempt, either, "Well done, you pressed the horn then the frogs", or "Remember, you need to press them just the way I say – first the horn, then the frogs" and the experimenter demonstrated. The task instructions were repeated if necessary.

In the test phase children were presented with lists varying in length from two to four items. The experimenter said the names of the items and the child had to press the picture locations on the grid. There were two lists of each length, and if a child failed both attempts at any list length the test stopped. For example, children who passed both two-item lists, but then failed both three-item lists, were not presented with four-item lists. However, children who failed only one three-item list did progress to four-item lists.

Our scoring took account of the length of the lists remembered and the dual requirements to recall the items and press them in the right order, with more credit being given for the latter. Children scored 0 if they could not press two items in the correct or incorrect order. Children who remembered at least one two-item list but in the incorrect order scored 1, children who remembered at least one two-item list in the correct order scored 2, children who remembered at least one three-item list scored 3 (incorrect order) and 4 (correct order), children who remembered at least one four-item list scored 5 (incorrect order) and 6 (correct order). Taking into account whether children remembered items in the correct order gave a wider spread of responses, as the vast majority of our participants could remember only two or three items, and remembering the list and pressing the corresponding locations on the grid, even in the wrong order, makes working memory demands. Furthermore, all children met the criteria for the points on the scale below the level they were awarded, for example, all children at level 5 had also remembered three items in the right order (level 4).

Counterfactual reasoning measures

Syllogisms. Children were engaged in a warm-up task to encourage them to use imagery (see Dias & Harris, 1988, 1990). They were asked to imagine a dog and then imagine that it was roller-skating. The experimenter said she was also imagining this. Children were asked whether the dog was roller-skating or walking. This confirmed that children were willing to imagine an unusual event. A total of 74% of children said that the dog they imagined was roller-skating. The experimenter responded, "Well done, he's roller-skating!" For the other children the experimenter encouraged them saying, "I'm imagining that my dog is roller-skating!" The experimenter said that the stories would "sound a bit funny, but let's pretend everything in the stories is true". There followed four trials (see Appendix). The experimenter read the premises, e.g., "All sheep are green. Jenny is a sheep." and asked the test question, "In the pretend story, is Jenny green or white?" Children scored 1 if they gave the logical counterfactual answer (maximum = 4). Check questions at the start confirmed that children knew the relevant real-world facts. A total of 80% of children answered all check questions correctly. Mistakes were corrected by the experimenter.

Conditionals: Causal chains. For this task (German & Nichols, 2003) children were presented with two short stories (see Appendix), each depicting a series of four events linked in a causal chain and illustrated with a row of four pictures. In each story the main character was happy, suffered a mishap, and ended up sad. Children were asked two check questions about the character's emotion at the beginning and the end of the story, and 96% of children answered all these questions correctly. A "short" counterfactual question required the child to infer what would have happened had the event depicted in picture 3 not occurred, and a "long" counterfactual question required them to infer the outcome had the event depicted in picture 1 not occurred. A filler question, e.g., "What colour is Nicholas' top?" was asked in between the counterfactual questions to separate them. We counterbalanced the order of test questions (children had either a short or a long question first on both stories). There was no difference in performance between these groups. Children scored 1 on each story if they gave the counterfactual answer (maximum = 2), i.e., they judged that had the event in question not occurred, the protagonist would have been happy.

Conditionals: Location change. In this test (Riggs et al., 1998) children heard two stories that were acted out with toys. In the first story Piglet was drawing a picture at a table. The wind blew the picture into a tree. Children were asked the counterfactual question, "What if the wind hadn't blown, where would the picture be, would it be on the table or in the tree?" In a second story Mum baked a cake as described in the introduction. Children scored 1 on each story if they gave the counterfactual answer (maximum = 2), i.e., that the object would have been in its original location.

RESULTS

We present our statistical analysis in three stages. First, we examined the raw data and present descriptive statistics. Second, we looked for relationships between our measures using correlations. Finally, we used multiple regression to test which measures predicted counterfactual thinking.

Descriptive statistics

Table 1 presents descriptive statistics for children on all tasks. As our data were not normally distributed (apart from BPVS scores) we used non-parametric analyses where appropriate. There were no gender differences on any measures (Mann-Whitney U tests: highest $Z = -.86$, lowest $p = .389$).

In German and Nichols' (2003) study, counterfactual questions about short causal chains were easier than those about long causal chains. We

TABLE 1
Frequencies and mean scores on measures

Measure	*Frequencies for counterfactual tasks*		
Conditionals (0–2) $N = 93$	0	1	2
Short Causal Chains	38	19	36
Long Causal Chains	27	27	39
Location Change	39	24	30
Syllogisms (0–4) $N = 91$	0–1	2	3–4
Syllogism	33	22	36
	Mean scores on individual difference measures (SD)		
BPVS	41.58 (10.35)		
Bear/Dragon (range 0–8)	5.09 (3.17)		
Black/White (range 0–16)	12.45 (4.16)		
Noisy Book (range 0–6)	2.85 (1.10)		
Counting & Labelling (range 0–4)	1.23 (1.66)		

failed to replicate this finding: a Wilcoxon signed ranks test showed no difference between short and long chains, $Z = -1.29$, $p = .199$. This is discussed further in the General Discussion (see also Beck, Riggs, & Gorniak, in press; Chan & Hahn, 2007).

We compared children's performance on our counterfactual measures to the expected distribution if they were guessing. Children's performance was different to chance on all measures: Short, $\chi^2(2, N = 93) = 32.6$, $p < .001$; Long, $\chi^2(2, N = 93) = 19.5$, $p < .001$; Location change, $\chi^2(2, N = 93) = 23.5$, $p < .001$; Syllogisms, $\chi^2(4, N = 91) = 102.9$, $p < .001$.

Correlations between measures

We used Spearman's rank correlations to examine the relations between our individual measures; these are presented in Table 2. There were strong correlations between many of the measures. This is expected in a developmental study such as this: children who are relatively mature and perform well on one measure are likely to do well on other measures, There were strong correlations between three of the counterfactual measures (all $\leq .001$). However, the long causal chain measure did not correlate with the other counterfactual measures (in fact, the only relationship was a negative correlation with the Stroop). Because we failed to find the reported difference between long and short causal chains, and because we found that long causal chains did not correlate with other measures, we do not include the long causal chain measure in our analysis any further, although we will return to it in the discussion. We inferred from the other correlations that

TABLE 2
Correlations between measures, Spearman's rho

	BPVS	*Bear/ Dragon*	*Stroop*	*Counting and Labelling*	*Noisy Book*	*Short causal chains*	*Long causal chains*	*Location change*	*False syllogisms*
Age (months)	.26*	.13	.32*	.15	.12	.08	−.05	.14	.05
BPVS	–	.44**	.52**	.47**	.21*	.34**	−.09	.43**	.33**
Bear/Dragon		–	.44**	.33**	.27**	.38**	−.06	.42**	.28**
Stroop			–	.34**	.24*	.21*	−.22*	.31**	.28**
Counting and Labelling				–	.29**	.26*	.01	.26*	.21*
Noisy Book					–	.11	.04	.09	.25*
Short causal chains						–	.10	.49**	.37**
Long causal chains							–	−.14	.06
Location change								–	.33**

*Significant at the .05 level (two-tailed). **Significant at the .01 level (two-tailed).

the short causal chain conditionals, location change conditionals, and false syllogisms share reasoning demands.

There were correlations between our counterfactual measures and the executive function measures. The two inhibitory control measures were related to the counterfactual measures (the relationship between the Stroop and the short causal chains was not significant but showed a positive trend). The pattern for the working memory measures was less consistent. Counting and Labelling, but not Noisy Book, correlated with the counterfactual measures. To test whether these relationships exist independently of each other and general development we used a regression analysis.

Data coding

The individual measures were not suitable for a regression analysis (data were very unevenly distributed or scored on inappropriately short scales). We recoded these measures into composite measures of counterfactual thinking, inhibitory control, and working memory.

The counterfactual measures were recorded on very short scales (0–2), so we combined the three related measures (short causal chains, location change, and false syllogisms) into a single composite counterfactual thinking score. Because the syllogisms score was on a scale from 0 to 4, whereas the two conditional tasks were each scored 0 to 2, we halved the syllogisms score before summing it with the two conditional counterfactual scores.

There were strong correlations between the two inhibitory measures and between the two working memory measures. We created two composite measures, one for inhibitory control and one for working memory. Using a composite based on more than one task should give greater validity, as we have two measures that have been designed to tap the same construct. We observed that the spread of scores across the scales on these tasks was distributed very irregularly. On each of the two inhibitory control measures, one third of the children gained the top score. We therefore categorised children into three groups: those with the top score, and two further groups created using a median split of the remaining children. Having done this for both measures, we summed the scores to get a composite inhibitory score. We used a similar strategy for the working memory tasks. Approximately one third of the children scored each of 2, 3, or 4 on the Noisy Book measure. We combined the five children scoring 0 or 1 with those scoring 2 and two children scoring 5 or 6 with those scoring 4, to obtain three broadly equal groups. Performance on the Counting and Labelling task was more varied but a large number of children (55 of 92) scored 0. Thus we categorised children into two groups: poor performers on this task (60%) and good performers who succeeded at least once (40%). We combined the two scores in to a composite working memory score.

Regression analysis

We used multiple regression to test whether our executive function measures predicted counterfactual thinking ability. Although our data were not normally distributed, violations of this assumption are relatively unimportant if the other assumptions, homoscedasticity, linearity, outliers, and collinearity, are met (Allison, 1999). One measure, age, showed heteroscedasticity and so it was omitted from the regression. We did not expect age to contribute to our model, as it did not correlate with the majority of measures. Verbal ability should provide a better indicator of general intellectual development than a simple maturation measure. There was one outlier identified on the language measure (a child aged 4;1 who scored 71 on the BPVS. This score is typical of a 7-year-old). This child's data were removed for the regression analysis.

We used three independent variables, BPVS, Inhibitory Control, and Working Memory to predict the score on counterfactual thinking. The model was significant, $R^2 = .22$, $F(3,\ 89) = 9.36$, $p < .001$. Although the model accounts for only 22% of the variance, this level is typical in similar developmental studies (see, e.g., Carlson, Moses, & Breton, 2002; Carlson, Moses, & Claxton, 2004). Inhibitory Control, $\beta = .26$, $t = 2.15$, $p = .034$, and BPVS, $\beta = .27$, $t = 2.30$, $p = .024$, were significant predictors. Working Memory was not a significant predictor ($p = .637$).

Contrary to the predictions in the literature we found no relation between counterfactual thinking and working memory. We further investigated this in two ways. First we entered each working memory measure into the regression analysis separately in case the measures were tapping different constructs. Neither predicted counterfactual thinking. Second we re-examined the coding of the Noisy Book task, which was based on a combined scale of number of items remembered and recall order. Reanalyses using scoring systems based on only item or order memory produced the same pattern of results.

GENERAL DISCUSSION

Three tasks—short causal chains, location change conditionals, and false syllogisms—shared a common element, despite their varying incidental task demands. What is common to both false syllogisms and counterfactual conditionals is the need to reason with something that is known to be false. In the syllogism tasks this is a statement that children know is false in the real world. In the conditionals children have to imagine a different version of the events they have just witnessed. In this sense, both syllogisms and conditionals involve thinking about a counterfactual world.

Despite this finding that three counterfactual measures tapped a shared ability, one task—long causal chains—was noticeably different. We failed to find the difference reported by German and Nichols between short and long causal chains. However, there are several reasons to think our findings are not an anomaly. First, another lab has failed to replicate the German and Nichols result (Chan & Hahn, 2007) and in four further experiments ($N = 192$), in which we tried to explore why long causal chains might be more difficult for children than short, we never replicated this original pattern (see Beck et al., in press). Second, German and Nichols used a between-participants design and a relatively small sample (13 children aged 3 and 13 aged 4 who answered short questions, and 13 children aged 3 and 15 aged 4 who answered long questions). We used a larger sample (47 children aged 3 and 44 aged 4) and a within-participants design, giving us more power to make comparisons between the two measures. Third, in our study the long causal chain measure did not correlate with any other measure. All our other measures, executive function and counterfactual alike, correlated strongly with BPVS score. This is to be expected from a battery of cognitive tasks. The validity of the long causal chains measure is called into question by the fact that it was the only measure not to correlate with the BPVS. This raises a concern that children may have been guessing answers to the long causal chain questions, or using some systematic strategy for answering the questions (for example, describing the content of the picture the experimenter pointed to). We do not address children's performance on long causal chains further here, but are cautious about it as a measure of counterfactual thinking.

Which general skills do children need to think counterfactually? We found that inhibitory control predicted performance on counterfactual thinking tasks. This supports the idea that 3- to 4-year-olds' difficulty with counterfactual thinking is in ignoring what they know to be true. In our tasks the to-be-ignored knowledge was a currently visible state of affairs (in the location change task), a mishap described in a story (causal chains), or general knowledge (false syllogisms). What children know to be true about the world competes with the thought of what might have been.

We found no evidence that working memory relates to developments in counterfactual thinking ability. To some extent this is surprising, as we anticipated a role for working memory based on our task analysis and other data: one needs to hold in mind both the actual and known to be false world when thinking counterfactually (see Byrne, 2005). However, recent work may explain the lack of any relation here. Beck, Robinson, Carroll, and Apperly (2006) argue that when younger children answer counterfactual conditionals correctly they are focusing *only* on the counterfactual world, and they do not recognise the relation between the counterfactual and actual world that both were once possible. As a consequence there are no additional working memory demands. Thinking about counterfactuals as possibilities—a development that happens at around 5 years—may make substantial demands on working memory, as one has to hold both possible worlds in mind.

It is important not to overlook the strong relationship between language and counterfactual thinking observed in this study. Our receptive vocabulary measure predicted performance on our counterfactual measure. This may reflect one of various roles language could play in counterfactual thinking. It might be that the BPVS measure is a proxy for general cognitive development, and this accounts for improvements in children's counterfactual thinking. Alternatively, the counterfactual reasoning measures may require substantial linguistic ability. Language is likely to be important to cue one to imagine what might have been. Future research should investigate whether specific linguistic developments might predict children's counterfactual thinking (as is the case with theory of mind, see de Villiers & Pyers, 2002; Smith, Apperly, & White, 2003) or whether general language development enables children to create more complex models of the world including counterfactuals.

It is possible that presenting the trials in a fixed order could have affected children's performance on particular tasks. For example, counterfactual syllogisms were always presented last, which might have enhanced performance due to training or reduced performance due to fatigue. We tried to minimise these effects by distributing the related tasks across two sessions, and there is no a priori reason why the order we presented would have led to effects that could have caused the apparent relationships between tasks. However, a counterbalanced comparison of tasks would complement our study and confirm whether these tasks differ in difficulty.

Conclusions

We found that three of the tasks used to measure children's emerging counterfactual thinking are strongly related. These tasks make different incidental demands, but share the need to reason with something known to be false.

Language ability accounted for significant variation in children's counterfactual thinking. Counterfactual thinking also makes demands on inhibitory control, as the child must ignore what is known to be true.

Manuscript received 22 October 2008
Revised manuscript received 31 March 2009
First published online 11 August 2009

REFERENCES

Allison, P. D. (1999). *Multiple regression: A primer*. Thousand Oaks, CA, USA: Sage.

Baird, A., & Fugelsang, J. (2004). The emergence of consequential thought: Evidence from neuroscience. *Philosophical Transactions of The Royal Society of London: Series B*, *359*, 1797–1804.

Beck, S. R., Riggs, K. J., & Gorniak, S. L. (in press). The effect of causal chain length on counterfactual conditional reasoning. *British Journal of Developmental Psychology*.

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.

Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development*, *72*(4), 1032–1053.

Carlson, S. M., Moses, L. J., & Breton, C. (2002). How specific is the relation between executive function and theory of mind? Contributions of inhibitory control and working memory. *Infant and Child Development*, *11*, 73–92.

Carlson, S. M., Moses, L. J., & Claxton, L. J. (2004) Individual differences in executive functioning and theory of mind: An investigation of inhibitory control and planning ability. *Journal of Experimental Child Psychology*, *87*, 299–319.

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, 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.

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.

Dunn, L. M., Dunn, L. M., Whetton, C., & Burley, J. (1997). *British Picture Vocabulary Scale*. (2nd ed.). London: NFER Nelson.

Espy, K. A., McDiarmid, M. M., Cwik, M. F., Stalets, M. M., Hamby, A., & Senn, T. E. (2004). The contribution of executive functions to emergent mathematic skills in preschool children. *Developmental Neuropsychology*, *26*, 1465–1486.

Flynn, E. (2007). The role of inhibitory control in false belief understanding. *Infant and Child Development*, *16*, 53–69.

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

Gerstadt, C. L., Hong, Y. J., & Diamond, A. (1994). The relationship between cognition and action - performance of children 3 ½ – 7 years old on a Stroop-like day-night test. *Cognition*, *53*, 129–153.

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.

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

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

Hughes, C. (1998). Executive function in preschoolers: Links with theory of mind and verbal ability. *British Journal of Developmental Psychology*, *16*, 233–253.

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.

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.

Markovits, H., & Doyon, C. (2004). Information processing and reasoning with premises that are empirically false: Interference, working memory, and processing speed. *Memory and Cognition*, *32*, 592–601.

Neys, W., & Everaerts, D. (2008). Developmental trends in everyday conditional reasoning: The retrieval and inhibition interplay. *Journal of Experimental Child Psychology*, *100*, 252–263.

Nolan, B. (1993). *Little Red Riding Hood rides again*. Racine, WI: Western Publishing Company, Inc.

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.

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., 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.

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. (2005). Inhibitory and working memory demands of the day–night task in children aged 3 to 11 years. *British Journal of Developmental Psychology*, *23*, 471–486.

Smith, M., Apperly, I. A., & White, V. (2003). False belief reasoning and the acquisition of relative clause sentences. *Child Development*, *74*(4), 1709–1719.

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.

APPENDIX

False syllogisms

All sheep are green
Jenny is a sheep
In the pretend story, is Jenny green or white?

All elephants are purple
Rosie is an elephant
In the pretend story, is Rosie purple or grey?

All horses eat pizza
Michael is a horse
In the pretend story, does Michael eat pizza or grass?

All cats chase crocodiles
Tom is a cat
In the pretend story, does Tom chase crocodiles or mice?

Causal chains

(The order of short and long test questions was counterbalanced.)

Story 1

1. Here is Mrs Rosy. She's just planted her new flower and she's very happy with it. She calls her husband from the house to come and have a look.
2. When Mr Rosy opens the door to come into the garden, the dog escapes from the kitchen.
3. The dog runs around the garden. Look he jumps on the flower and squashes it!
4. Now the flower is all flat, and Mrs Rosy is sad.

Check 1: Just now, is Mrs. Rosy happy or sad?
Check 2: Right at the beginning, was Mrs. Rosy happy or sad?
Short test question (picture 3): What if the dog hadn't squashed the flower, would Mrs Rosy be happy or sad?
Filler: What colour is Mum's top?
Long test question (picture 1): What if Mrs. Rosy hadn't called her husband, would Mrs Rosy be happy or sad?

Story 2

1. Nicholas is playing with his balloon in the garden. He's very happy. Mum calls him into the house to have his snack.
2. Nicholas gives the balloon to his friend to look after. His friend plays in the garden.

3. The friend runs too close to the rosebush. He falls into the bush and the balloon pops.
4. Now the balloon is popped, and Nicholas is sad.

Check 1: Just now, is Nicholas happy or sad?
Check 2: Right at the beginning, was Nicholas happy or sad?
Short test question (picture 3): What if his friend hadn't fallen in the rosebush. Would Nicholas be happy or sad?
Filler: What colour is Nicholas' top?
Long test question (picture 1): What if Mum hadn't called Nicholas in for a snack. Would he be happy or sad?