Developmental Science 6:5 (2003), pp 514–523

PAPER

Children's counterfactual inferences about long and short causal chains

Tim P. German[1] and Shaun Nichols[2]

1. Department of Psychology, University of California at Santa Barbara, USA
2. Department of Philosophy, College of Charleston, USA

Abstract

Recent findings on counterfactual reasoning in children have led to the claim that children's developing capacities in the domain of 'theory of mind' might reflect the emergence of the ability to engage in counterfactual thinking over the preschool period (e.g. Riggs, Peterson, Robinson & Mitchell, 1998). In the study reported here, groups of 3- and 4-year old children were presented with stories describing causal chains of several events, and asked counterfactual thinking tasks involving changes to different points in the chain. The ability to draw successful counterfactual inferences depended strongly on the inferential length of the problem, and the age of the children; while 3-year-olds performed above chance on short inference counterfactuals, they performed below chance on problems involving longer inference chains. Four-year-old children were above chance on all problems. Moreover, it was found that while success on longer chain inference problems was significantly correlated with the ability to pass tests of standard false belief, there was no such relationship for short inference problems, which were significantly easier than false belief problems. These results are discussed in terms of the developmental relationships between causal knowledge, counterfactual thinking and calculating the contents of mental states.

Introduction

The ability to think about counterfactual situations – situations that might have happened, but did not in fact happen – is implicated in a wide variety of cognitive capacities, both in research in adult cognitive psychology (e.g. Kahneman & Tversky, 1982) and in research on the developing cognitive capacities of children. It has been linked to the understanding of causation (e.g. Harris, German & Mills, 1996; German, 1999), the capacity for pretense (e.g. Harris, 1992; Leslie, 1987; Nichols & Stich, forthcoming) and the understanding of permission rules (e.g. Harris & Núnez, 1996; Harris, 2000).

In several recent publications, Riggs, Peterson and colleagues (Riggs *et al.*, 1998; Peterson & Riggs, 1999; Peterson & Bowler, 2000; Riggs & Peterson, 2000) have argued that the broad ability to handle counterfactual thinking is limited in the preschool period, and that this limitation explains the difficulty young children have in tasks assessing 'theory of mind' (e.g. the false belief task, Wimmer & Perner, 1983; Baron-Cohen, Leslie & Frith, 1985). In a series of four studies, Riggs *et al.* found that children aged $3^1/_2$ to $4^1/_2$ years had considerable difficulty with counterfactual thinking tasks. In one task for example, children were told a story about a man who goes to bed because of illness, but is then called to put out a fire at the Post Office. Children were asked where the man would be if there had been no fire. Less than 50% of the 3–4-year-old subjects succeeded on this task (Riggs *et al.*, 1998, Experiment 1). Moreover, as the authors predicted, performance on the counterfactual task was significantly related to performance on a false belief version of the same task. This was the pattern that emerged over a series of four studies (see also Robinson & Beck, 2000).

As Riggs and colleagues interpret their findings, the results have dramatic implications. First, the findings seem to challenge the claim that children have a precocious capacity to engage in counterfactual reasoning and that young children often assess the cause of a past event by considering counterfactuals (e.g. Harris *et al.*, 1996; German, 1999). Second, and more strikingly,

Address for correspondence: Tim German, Department of Psychology, University of California, Santa Barbara, CA 93106-9660, USA; e-mail: german@psych.ucsb.edu

Riggs *et al.* (1998) and Peterson and Riggs (1999) suggest that the findings provide the basis for a novel explanation of performance on the false belief task. On their view, children fail classic false belief tasks because they cannot perform the counterfactual reasoning that is required (e.g. Peterson & Riggs, 1999, p. 80). Thus, the work of Riggs, Peterson and their collaborators seems to demand a radical departure from the major theorizing on the false belief task and the emerging work on causal reasoning.

Despite the exciting prospect offered by Riggs and colleagues, there is an immediate empirical problem. For there are results in the literature showing that young children *succeed* in counterfactual tasks. Harris *et al.* (1996) presented several experiments suggesting quite precocious counterfactual thinking among 3-year-olds in tasks involving counterfactual thinking about simple causal events. In one study, 3-year-old children were presented with stories of simple causal sequences, acted out with props. For example, one story described a girl who came home and forgot to wipe her shoes before going into the kitchen, resulting in muddy footprints across the kitchen floor. Children were asked whether the floor would still be dirty if the character had taken her shoes off. The results showed that 3-year-olds performed above chance (75%), nearly as well as the 4-year-old group (84%). Harris *et al.* suggest that from the preschool years, children's precocious counterfactual thinking might be a critical part of their ability to come to causal judgments about events (see also German, 1999; Harris, 2000).

Recently, theorists have tried to explain the discrepancy between performance in Riggs *et al.* and in Harris *et al.* Harris and Leevers (2000) suggest that the critical difference might have to do with the fact that Harris *et al.* (1996) used concrete causal stories involving minor mishaps, which might facilitate counterfactual thinking through provoking children into considering how the outcomes might have been prevented. By contrast, Riggs *et al.* (1998) used neutral events, some of which involved counterfactual thinking about rather abstract situations such as change in an object's location, shape or color, rather than deviations from concrete causal scenarios. Robinson and Beck (2000) suggest instead that the difference between the Harris *et al.* studies and those of Riggs *et al*. (1998) is to be found in the fact that in the Riggs *et al*. tasks children 'had to identify what the counterfactual situation would be had some different action been taken' whereas in the Harris *et al*. tasks, the child can 'just judge that the current situation would have been avoided, or how it could have been avoided' (Robinson & Beck, 2000, p. 102).

We offer an alternative explanation for the discrepancy. We suggest that preschool children's difficulty in drawing counterfactual inferences does not stem from a problem in counterfactual reasoning *per se*, but rather from *the complexity of the inferences* required by the various tasks used. In particular, evidence will be provided for the view that a powerful determinant of children's performance is the length of the causal chain between the premise to be altered and the counterfactual outcome to which they must reason. This hypothesis is derived from analysis of the particular tasks used by the different authors. First, all the stories used in Harris *et al*. (1996) involved simple causal stories. In each case there was a starting situation (e.g. clean floor), an event occurred (e.g. Carol walked across the floor with muddy shoes), and the situation was changed (e.g. floor becomes dirty). The questions asked about the situation that would obtain had the event not occurred, requiring an inference something like the following: REMOVE SHOES – NO MUDDY FOOTPRINTS. Now consider one of the stories used by Riggs *et al*. (1998). In their 'Post Office' story, Peter is in a house but is feeling unwell and goes to bed. The phone rings and the man from the Post Office asks Peter if he will come and help put out a fire. Peter gets out of bed and goes to the Post Office. Children were asked where Peter would be if there had been no fire. Here children must infer something like: NO FIRE – NO PHONE CALL – PETER NOT GO TO POST OFFICE – PETER IN BED. This task appears to require a significantly longer chain of inferences than the Harris *et al*. tasks do. This, we suspect, explains why children tended to do so much worse on the tasks in Riggs *et al*. (1998). However, it's important to note that one of the tasks in Riggs *et al*. does not require a long chain of inferences. In the 'picture' story, a picture is blown by the wind from a table into a tree and the child is asked where the picture would be if the wind had not blown. Here the inference required of the children is simpler: NO WIND – PICTURE NOT BLOW OFF TABLE – PICTURE ON TABLE. But, to the surprise of the researchers, in the experiment using this task, the children actually performed better on the counterfactual cases than on the false belief task (Riggs *et al*., 1998, p. 79). Hence, this further suggests that the length of inference might be the crucial factor that explains the discrepancy between Harris *et al*. and Riggs *et al*. Unfortunately, further analysis is not possible because Riggs *et al*. do not provide success rates on a task-by-task basis, nor do they provide precise details of their 'chocolate' story or their 'shop' story.

In the experiment to be presented here, children's ability to draw counterfactual inferences from causal chains of different lengths will be assessed. More precisely, we will compare performance in tasks that differ in the length of inference required between the antecedent event changed, and the outcome that follows.

Experiment

All children were presented with stories involving a series of four events linked in a causal chain. For example, children were told about Mrs Rosy, who is very happy with the flower she has planted in her garden, and calls her husband to come and see. When her husband opens the kitchen door, the dog escapes from the kitchen. The dog runs around the garden and jumps on the flower, squashing it. Mrs Rosy is very sad because the flower is squashed. Different groups of children were asked to reason about the consequences ensuing from changes to different points in this story. For example, in the 'long chain' inference condition, they were asked, what if Mrs Rosy hadn't called her husband from the house, would she be happy or sad? These children were required to reason that if she hadn't called her husband then the dog wouldn't have escaped, wouldn't have run in the garden, and wouldn't have squashed the flower, to the conclusion that Mrs Rosy would be happy. 'Medium chain' inferences were created by asking children to imagine if the dog hadn't escaped from the house (then he wouldn't have run in the garden, wouldn't have squashed the flower, so Mrs Rosy would be happy). Finally, the simplest tasks involved 'short chain' inferences, asking children about whether Mrs Rosy would be happy or sad if the dog had not squashed the flower.

In order to assess the relationship between the ability to draw such counterfactual inferences and the ability to reason about mental states, children in this experiment were also presented with two standard measures of false belief reasoning. Note that Riggs *et al*. assessed false belief performance with tasks modeled very closely on their counterfactual inference tasks. Using closely matched tasks can be a strength in studies of cognitive development if one shows different, or unrelated performance, because such differences can then be attributed to deeper differences in the conceptual competence or in the resources underlying the two tasks. For example, the structural match between the false photograph task (e.g. Zaitchik, 1990) and the false belief task is a strength when one comes to interpret the very different patterns of performance on these two tasks in individuals with autism (see e.g. Leslie & Thaiss, 1992; Charman & Baron-Cohen, 1995). However, when tasks are highly correlated, drawing the conclusion that they share deep conceptual competence is in fact more persuasive if their superficial structures are quite *different*. If the task structures are similar then one never knows whether any correlation results from the alleged 'deep' conceptual and/or resource similarities, or more superficial features that the tasks also have in common. In the current experiment we therefore used standard measures of false belief – the 'Sally-Anne' task (Baron-Cohen *et al*., 1985), and the 'Smarties' task (Gopnik & Astington, 1988; Perner, Leekham & Wimmer, 1987) – in order to be sure that any relationship between false belief and counterfactual thinking tasks could not be attributed to superficial structural similarities.[1]

From the work of Harris *et al*., we predicted that children would find shorter counterfactual inference chains easier. The short inference problems were designed to be similar to the questions posed about the simple scenarios used in Harris *et al.* (1996, Experiment 1; see also Riggs *et al*., 1998, Experiment 2; Robinson & Beck, 2000, study 7). However, as the length of the inferential chain in the counterfactual tasks increases, we would expect that children would begin to have more difficulty, as suggested by the work of Riggs *et al*. (1998).

Method

Participants

Eighty-three children participated. Forty-one 3-year-olds were assigned randomly to either the short inference condition (8 boys, 5 girls, mean age 3;5, range 3;1 to 3;11), medium inference condition (5 boys, 10 girls, mean age 3;4, range 3;0 to 3;10) or the long inference condition (6 boys, 7 girls, mean age 3;5, range 3;0 to 3;10). Forty-two 4-year-olds were assigned randomly to either the short inference condition (9 boys, 4 girls, mean age 4;8, range 4;2 to 5;0), medium inference condition (9 boys, 5 girls, mean age 4;8, range 4;1 to 4;11) or the long inference condition (7 boys, 8 girls, mean age 4;8, range 4;1 to 5;0). One further child was excluded owing to failure to complete the procedure. The children were recruited from nursery schools in the region of Athens, Greece. All children had Greek as a first language.

Tasks and materials

Two stories were created for the study. They were presented to the children using four colored picture cards (12 cm × 10 cm). Each story involved four events, one linked to each of the pictures. The pictures for the

[1] The Smarties task for 'self' has been the focus of the most extensive research on children's ability to handle beliefs caused by deceptive appearances (e.g. Barreau & Morton, 1999; Freeman & Lacohée, 1995; Mitchell & Lacohée, 1991). We chose this version here because though it is highly correlated with the version for 'other' (Gopnik & Astington, 1988), on the face of it, it is somewhat *less* similar in structure to the Sally-Anne task precisely because the latter asks about another's false belief. This makes it especially suited for the current experiment given the logic outlined in the previous paragraph.

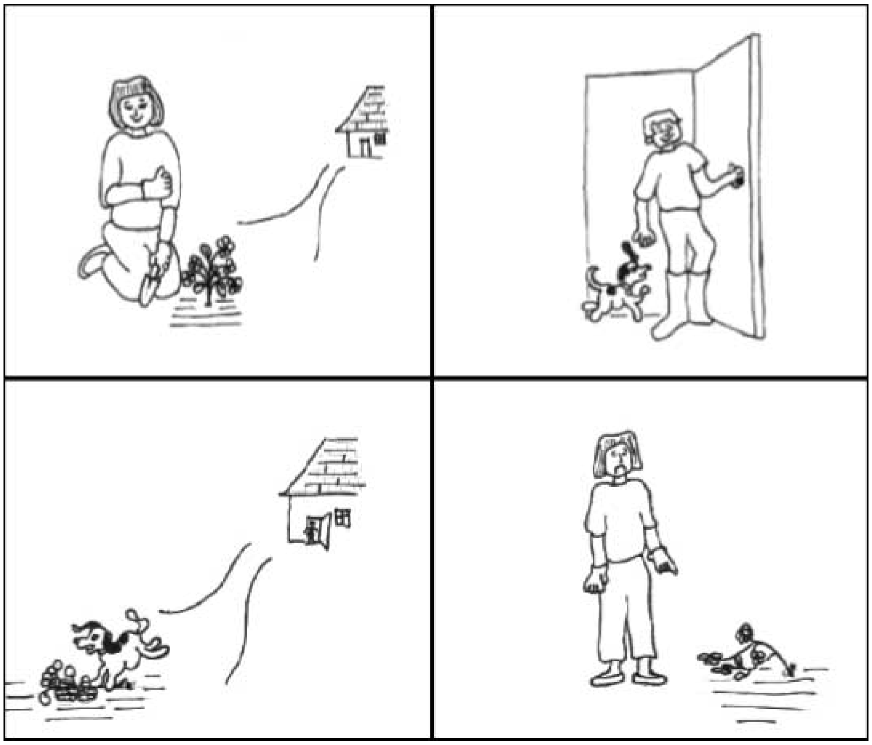

Figure 1 *Pictures accompanying the 'flower' story in the experiment.*

'flower' story are shown in Figure 1, and described below, in each of its three versions. The other story details appear in the Appendix.

> [Picture 1 top left]: '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.' [Picture 2 top right]: 'When Mr Rosy opens the door to come into the garden, the dog escapes from the kitchen.' [Picture 3 bottom left]: 'The dog runs around the garden. Look he jumps on the flower and squashes it!' [Picture 4 bottom right]: 'Now the flower is all flat, and Mrs Rosy is sad.'

All children were asked the 'Now control question': 'Just now, is Mrs Rosy happy or sad?', and the 'Before control question': 'Right at the beginning, was Mrs Rosy happy or sad?' They were then asked a counterfactual test question depending on the condition to which they had been assigned:

> *Short inference condition*: [Point at picture 3] 'What if the dog hadn't squashed the flower, would Mrs Rosy be happy or sad?'
> *Medium inference condition*: [Point at picture 2] 'What if the dog hadn't escaped from the house, would Mrs Rosy be happy or sad?'
> *Long inference condition*: [Point at picture 1] 'What if Mrs Rosy hadn't called her husband, would Mrs Rosy be happy or sad?'

Materials and procedure for the Sally-Ann and Smarties tasks were as found in the standard descriptions (Baron-Cohen *et al.*, 1985; Gopnik & Astington, 1988).

For all tasks, the stories were generated initially in English, before being translated into Greek. The Greek versions were then back-translated by a non-psychologically trained native Greek speaker, and minor changes were made to the initial translations to arrive at versions suitable for preschool children. The changes to the stories concerned only minor wording issues (e.g. 'squashes' was used rather than the initial 'flattens'). The back-translations are presented here (and in the Appendix).

Procedure

Children were tested individually in a quiet corner of the classroom by a female experimenter with whom they were familiar. Children were presented with the two counterfactual stories first, always in the order 'flower', 'balloon'. Following this, they received the two false belief tasks, in the order 'Sally-Ann', 'Smarties'. Each counterfactual story was presented one picture at a time, with the pictures laid out in sequence on the floor. Children were asked the two control questions, followed by a counterfactual question according to the condition. The experimenter pointed to the relevant picture when

Table 1 *Percentage correct for each task according to age and condition and mean counterfactual and false belief score (from 2; SD in parentheses)*

	Counterfactual tasks			False belief tasks		
	Flower	Balloon	Score (SD)	Sally-Ann	Smarties	Score (SD)
Short inference						
3-year-olds ($N = 13$)	69	69	1.38 (0.96)	31	23	0.54 (0.88)
4-year-olds ($N = 13$)	100	100	2.00 (0.00)	62	76	1.38 (0.77)
Medium inference						
3-year-olds ($N = 14$)	20	33	0.53 (0.83)	20	33	0.53 (0.83)
4-year-olds ($N = 15$)	86	86	1.71 (0.72)	64	57	1.21 (0.98)
Long inference						
3-year-olds ($N = 13$)	15	15	0.31 (0.75)	15	23	0.38 (0.77)
4-year-olds ($N = 15$)	67	67	1.33 (0.98)	73	80	1.53 (0.74)

asking the counterfactual question. After the four tasks had been completed, children were rewarded with a sticker. As noted in the participant section, one child failed repeated control questions and was excluded from further analysis.

Results

Children were scored correct on the counterfactual stories if they answered that the character in question (e.g. 'Mrs Rosy') was happy. Preliminary analysis revealed that there was no overall difference between the two stories ('flower' story, $M = .59$, SD = .49; 'balloon' story $M = .61$, SD = .49, $t(82) = -1.43$, NS; McNemar, binomial, NS). Moreover, the two counterfactual stories correlated highly and significantly ($r_\phi = .95$, $\chi^2_{[corr]N=83,\, df=1} = 71.13$, $p < .0001$, two-tailed), and so the scores from the two stories were added to create an aggregate 'counterfactual' score for each child. Children were scored correct on the Sally-Anne task if they pointed to or indicated verbally the marble's original location (basket). Children were scored correct on the Smarties task if they reported their own earlier belief (e.g. Smarties), and reality (e.g. pencil), correctly. Preliminary analysis again revealed that there was no overall difference between performance on the Sally-Anne and Smarties tasks (Sally-Anne, $M = .46$, SD = .50; Smarties, $M = .49$, SD = .50, $t(82) = -1.16$, NS; McNemar, binomial, NS), and once again the tasks were highly and significantly correlated ($r_\phi = .714$, $\chi^2_{[corr]N=29,\, df=1} = 39.47$, $p < .0001$, two-tailed), therefore they were summed to create an aggregate 'false belief' score for each child.[2]

The mean counterfactual and false belief scores, along with the percentage of children correct for each task in each condition at each age, appear in Table 1.

These aggregate counterfactual scores were submitted to a 2 (Age group: 3- versus 4-year-olds) × 3 (condition: short chain versus medium chain versus long chain inferences) ANOVA revealing main effects of age ($F_{1,\,83} = 29.47$, $p < .0001$) and condition ($F_{2,\,83} = 8.45$, $p < .0001$), but no interaction ($F < 1$). Post-hoc comparisons revealed that children performed better at short chain inferences than at either longer chain condition (Tukey's HSD, ps < .019, .001, respectively) but that the two longer chain conditions did not differ (Tukey's HSD, NS).

The relationship with performance on false belief was assessed via partial correlations controlling for age, for each condition. This analysis revealed a strong and highly significant correlation between both medium chain counterfactual inferences and false belief ($r_{p\,26} = .68$, $p < .0001$, two-tailed) and long chain counterfactual inferences and false belief ($r_{p\,25} = .60$, $p < .001$, two-tailed). There was a much weaker non-significant correlation between the short chain counterfactual inferences and false belief ($r_{p\,23} = .36$, $p = .08$, two-tailed).

Owing to the dichotomous nature of the underlying distributions, the results were also analyzed non-parametrically. For both counterfactual inferences and false belief reasoning a criterion of 2 out of 2 was set for a 'pass'. The probability of 'passing' either kind of task by chance was thus .25. Children in the short chain condition in both age groups performed significantly above chance (binomial theorem, where $N = 13$, p[success] = .25, critical frequency for $p < .01 = 7$). However, in both the medium chain condition and longer chain condition, the 3-year-olds did not perform better than chance while the 4-year-olds did (binomial theorem).

The percentage of children 'passing' each task in each condition, according to age, also appears in Table 1.

[2] These analyses also show that the consistent ordering of tasks within each block had no effect on children's performance.

These results show that while more 4-year-old children passed both tasks than did 3-year-old children in the medium and long chain counterfactual inference conditions ($\chi^2_{[corr]N=29, df=1} = 13.57$, $p < .0001$, two-tailed for medium chain problems; $\chi^2_{[corr]N=29, df=1} = 10.03$, $p < .001$, two-tailed for long chain problems), the difference for short chain problems fell short of significance ($\chi^2_{[corr]N=26, df=1} = 2.65$, $p > .1$, two-tailed). Turning to the effect of condition, and collapsing medium and long chain inference problems to meet the minimum frequency requirement for chi-squared analysis, there was also a significant effect of condition in the 3-year-old group and a marginally significant effect among the 4-year-olds. Children performed better on short chain counterfactual inference problems than on either of the two longer chain problems ($\chi^2_{[corr]N=26, df=1} = 8.26$, $p <$ .001, two-tailed and $\chi^2_{[corr]N=26, df=1} = 3.12$, $p < .08$, two-tailed, for 3- and 4-year-old groups, respectively).

Finally, children performed better on the short chain counterfactual inference tasks than they did at false belief (McNemar, binomial, $N = 12$, $k = 0$, $p < .00001$), while performance on medium and long chain inference problems was in line with false belief (McNemar, binomials, NS).[3]

Non-parametric correlations confirmed the picture provided by the parametric analysis in revealing strong significant correlations between both medium and long chain counterfactual inferences and false belief ($r_\phi = .76$, $\chi^2_{[corr]N=29, df=1} = 13.57$, $p < .0001$, two-tailed; $r_\phi = .86$, $\chi^2_{[corr]N=29, df=1} = 17.09$, $p < .0001$, two-tailed for medium and long chain problems, respectively). Again, there was a lower, non-significant association between the short chain counterfactual inferences and false belief ($r_\phi = .34$, $\chi^2_{[corr]N=26, df=1} = 1.35$, NS).

Discussion

The results of this experiment were clear. Children's ability to draw counterfactual inferences was strongly affected by the complexity of the inference task that they faced. Children required to infer the counterfactual outcome resulting from a change to an event several steps away in a causal chain performed more poorly than children who were required to reason from short causal chains. In general, 3-year-old children failed to draw counterfactual inferences for medium and long causal chains, while succeeding at short chain problems. This result replicates the findings of Harris *et al.* (1996, Experiment 1) suggesting that preschool children have no particular difficulty in drawing counterfactual conclusions from false premises. The 4-year-old group performed at ceiling on the counterfactual inferences from short chains, and performed better than chance at problems based on medium and long chains.

Thus, the results help confirm our explanation for the differences between children's performance in Harris *et al.* and in Riggs *et al.* The tasks in Harris *et al.* involved short counterfactual inferences, whereas the difficult tasks in Riggs *et al.* involved longer chain counterfactual inferences. The current experiment demonstrates that 3-year-olds in fact do well on short chain inferences, but poorly on longer chain inferences.

In addition, the differences manifested in the experiment cannot be explained by the alternative proposals. Harris and Leevers (2000) propose that children perform better in Harris *et al.* (1996) than in Riggs *et al.* (1998) because the former used *concrete* stories with *negative* outcomes whereas Riggs *et al.* used several abstract stories without negative outcomes. However, since *all* the stories used in this experiment were concrete and had minor mishaps as the final outcome, the difference between short chain and longer chain problems cannot be explained in terms of concreteness or negative outcomes.

The alternative explanation offered by Robinson and Beck is that 3-year-olds do well in the Harris *et al.* tasks because all they need do 'is judge that the current situation would have been avoided'. However, in the current experiment, the children don't merely report how the current situation would have been avoided, they needed in each condition to identify what the counterfactual situation would have been – that Mrs Rosy would have been happy.

The results of the experiment reported here also cast new light on the relationship between counterfactual thinking and reasoning about false belief. First, we found that children performed significantly better on short chain counterfactual tasks than on the false belief tasks. This suggests that young children's difficulty with false belief cannot be attributed to a broad inability to engage in counterfactual reasoning. Indeed our studies show that young children are quite capable of the kind of counterfactual reasoning that Peterson and Riggs (1999) claim is required by the false belief task. Interestingly, however, counterfactual thinking problems involving medium and long chain inferences correlated highly and significantly with performance on standard tests of false belief, even when age was taken into account. Furthermore, the correlations demonstrated here between medium and long chain counterfactual inference problems and false belief are based on 'standard' tests of

[3] Note that the finding that children performed better on the counterfactual problems only for short inference problems rules out the possibility that there was an overall tendency for children to perform differently on the block of tasks performed first.

false belief understanding (e.g. Baron-Cohen *et al.*, 1985; Perner *et al.*, 1987). This is an advance over the correlations shown in Riggs *et al.* (1998) because in that case the false belief measures were generated from questions based on exactly the same stories as the counterfactual measures. In the current study there were few superficial similarities between the 'Sally-Anne' and 'Smarties' stories used to assess false belief understanding and our picture-based counterfactual thinking tasks. This suggests that the correlations observed between the medium and long chain counterfactual inference tasks and false belief go beyond similarity at the level of superficial task structure.

How might one account for the relationship between counterfactual thinking and reasoning about false belief, if the difficulty with false belief cannot be attributed to a broad deficiency in counterfactual reasoning, as suggested by the present results? One possibility is that the medium and long chain inferences actually require an importantly different *kind* of counterfactual reasoning than the short chain inferences, and that this more complex kind of counterfactual thinking is also implicated in false belief reasoning.[4] This alternative has not been developed, but it offers an interesting approach to the issue. It is clear, however, that such an approach would not solve the problem that the present data pose for the position advanced by Riggs and Peterson. For the kind of counterfactual reasoning that they claim is implicated in false belief tasks is precisely the kind of *short chain* counterfactual inference that we test for in this study. The task analysis offered by Peterson and Riggs (1999) indicates that on the Maxi task, for instance, the crucial counterfactual reasoning is: 'If the chocolate was not moved to the cupboard (from the fridge), it would be in the fridge' (Peterson & Riggs, 1999, pp. 90–91). This seems to be exactly the kind of short chain counterfactual inference that young children succeeded at in the experiment reported here.

Although it is possible that there is some, as yet unarticulated, principled distinction between short chain counterfactuals and long chain counterfactuals, we are inclined to think that variance in the ability to generate counterfactual inferences is itself not causing the correlation with performance in false belief, but rather that the correlation results from similar information processing demands for medium and long chain counterfactual reasoning and false belief reasoning. Although it would be premature to settle on any particular processing story, we want to sketch out two ways that the correlation might be explained by appealing to certain kinds of shared information processing demands.

[4] This possibility was suggested by an anonymous referee.

There has been much recent interest in the idea that processing demands that go under the rather loosely specified term 'executive function' (e.g. Luria, 1966; Lezak, 1983) might be involved in the solution of false belief tasks. The notion of executive function has been developed in different ways, and it has been applied to false belief reasoning in different ways too (cf. Carlson, Moses & Hix, 1998; Gordon & Olson, 1998). The counterfactual inference problems used here actually seem ideal candidates to be considered tasks that make the 'classic' executive demands of inhibiting a pre-potent response while keeping relevant information in mind (see e.g. Russell, Saltmarsh & Hill, 1999).

A number of researchers within the executive function approach have focused on the role of *inhibitory* processes and suggested that false belief reasoning places heavy demands on inhibitory processes, since success on the false belief tasks requires subjects to inhibit a pre-potent response to attribute a true belief content (see e.g. Carlson *et al.*, 1998; Carlson & Moses, 2001; German & Leslie, 2000; Leslie & Pollizi, 1998). Counterfactual reasoning presumably also places heavy demands on inhibitory processes, and the correlation between reasoning about false beliefs and long chain counterfactuals might be attributed to these common factors. One *prima facie* problem for this kind of account as an explanation for the present results is that at a superficial level, *all* the counterfactual tasks employed here require that the child generate a response that is different from the 'current state of affairs'. If all the tasks require inhibitory processing to the same degree, then the account cannot explain the difference in performance observed between short inference problems and medium/long inference problems (cf. Robinson & Beck, 2000, p. 105).

However, one way that the medium and long chain counterfactual problems might place additional demands over the short chain tasks is in the need for maintaining the false proposition through further stages of inferential processing. Note that for each event in the chain, the outcome of mutating that event results in a proposition that differs from the sequence of events that actually happened. For example, if one can inhibit the salient fact that the dog escaped from the house, one can infer that the dog did not run in the garden. Note further that this proposition (*the dog did not run in the garden*) itself differs from the actual sequence of events, and further inhibitory resources might be required to prevent the true proposition (*the dog ran in the garden*) being entered into the next sequence of the inference, which is necessary in order to reach the proper conclusion – *Mrs Rosy would be happy*. The longer inference tasks therefore can be characterized as requiring either multiple inhibitory processes to be deployed as the inferential complexity increases, or for an inhibitory process to be maintained over time. At any point, insufficient inhibitory resources might result in the system ‘snapping back’ to the default state, and the true state of affairs (rather than a counterfactual premise) being entered into a given step in the process. If this kind of analysis is right, then the documented difficulty that children have with inhibitory processing might account for poor performance both in the medium and long chain counterfactual inference tasks and false belief problems, where the powerful default belief content must be inhibited and an alternative content generated and selected (e.g. Carlson & Moses, 2001; German & Leslie, 2000; Leslie & Pollizi, 1998). Interestingly, Leslie and Pollizi (1998) pursue the proposal that the capacity to handle multiple interacting inhibitory processes might underlie difficulties in belief desire reasoning that persist beyond the 3–5-year age range studied predominantly in the theory of mind literature.

An alternative information processing explanation is that the longer chain inferences require substantially greater working memory capacity than short chain inferences. For example, it might be the case that the entire chain of events must be stored by the child during the presentation of the problem and the effects of mutating one of the antecedent events must be generated from processing the stored representation of the event sequence in working memory. If the entire chain is not represented adequately, then the inference process could fail to result in a response, and the only available response would be the default reality-based answer. This kind of approach has been used to explain how increases in the capacity to ‘hold in mind’ might account for the development of theory of mind concepts (Davis & Pratt, 1995), and there have been correlations reported between children’s developing capacity for holding information in mind and performance on certain false belief tasks (e.g. Gordon & Olson, 1998).[5] Thus, one might maintain that young children’s difficulty with holding information in mind might explain the correlation between performance on false belief tasks and longer chain counterfactual reasoning.

[5] Note that the Davis and Pratt (1995) proposal concerns the development of the capacity for holding in mind as an explanation for the acquisition of *mental state concepts themselves*. It is therefore somewhat different from information processing proposals suggesting an early conceptual competence that is gradually more efficiently expressed, as inhibitory and other executive type processes develop (e.g. German & Leslie, 2000; Leslie, 1994; Leslie & Thaiss, 1992). Moses (2001) distinguishes these two types of proposal as ‘emergence’ versus ‘expression’ accounts of the relation between executive processes and ‘theory of mind’, favoring the latter proposal (e.g. Moses, 2001; Carlson & Moses, 2001).

If this latter kind of proposal is correct, then we can predict that correlations might emerge between the medium and long inference tasks we present here and the measures of working memory shown to correlate with false belief mentioned above. More directly, there should be a relationship between the children’s ability to answer correctly the longer counterfactual inference problems and their ability to demonstrate recall for the inference structure of the story. This could be assessed via tasks requiring children to order the pictures in sequence or retell the story with or without the pictures as cues (e.g. Baron-Cohen, Leslie & Frith, 1986). If the former, inhibitory proposal is correct, then correlations might be expected to emerge between the counterfactual inference tasks used here and other inhibitory measures of executive function (e.g. those employed by Carlson & Moses, 2001; Gerstadt, Hong & Diamond, 1994). On this account, we might make the prediction that increasing the length or complexity of the inference chain still further might lead to poorer performance among children capable of solving two- and three-step inferences and standard false belief tasks, and further that these longer chain counterfactual tasks would correlate with false belief tasks that have been proposed to place even higher inhibitory demands on children (e.g. Leslie & Pollizi, 1998; Leslie, German & Pollizi, in preparation).

We offer the foregoing merely as two possible explanations for the correlation between false belief reasoning and multi-step counterfactual reasoning. Moreover, the accounts need not be mutually exclusive. There are proposals that children’s ability to coordinate inhibitory and working memory capacities, e.g. to remember certain kinds of information (a causal chain) and exercise inhibitory control over that information, might underlie improving performance on certain executive tasks across the preschool and early school age years (e.g. Gerstadt *et al.*, 1994). However, research on ‘executive function’ in adult humans with and without brain injuries is beginning to provide a picture of interrelated processes that is far from simple (see e.g. Burgess, 2000; Rabbitt, 2000), so it is important to be cautious. Nonetheless, we suspect that information processing approaches provide the most promising way to explain the striking correlation between reasoning about long chain counterfactual problems and reasoning about false belief.

Appendix

Details of the Balloon story used in the experiment

[PICTURE 1: Shows boy playing in garden with balloon. He is smiling. Mother is pictured to one side]

'Nicolas is playing with his balloon in the garden. He's very happy. His mother calls him into the house to have his snack.'
[PICTURE 2: Shows boy handing the balloon to another boy in the garden] 'Nicolas gives the balloon to his friend to look after. His friend plays in the garden.'
[PICTURE 3: Shows boy falling into rosebush, balloon popping] 'The friend runs too close to the rosebush. He falls into the bush and the balloon pops.'
[PICTURE 4: Shows popped balloon on rosebush, Nicolas pointing with sad expression] 'Now the balloon is popped, and Nicolas is sad.'

Questions

Now control question: Just now, is Nicolas happy or sad?
Before control question: Right at the beginning, was Nicolas happy or sad?
Short inference counterfactual question: What if his friend hadn't fallen in the rosebush. Would Nicolas be happy or sad?
Medium inference counterfactual question: What if Nicolas hadn't given the balloon to his friend, would he be happy or sad?
Long inference counterfactual question: What if Mother hadn't called Nicolas in for snack. Would he be happy or sad?

Acknowledgements

We would like to thank Eleni Chalari, for help collecting the data presented here, and the teachers, parents and children of First Nursery School, Iraklio, Iliofeggaro Nursery School, Peuki and Melothia Nursery School, Amaroussion.

References

Baron-Cohen, S., Leslie, A.M., & Frith, U. (1985). Does the autistic child have a 'theory of mind'? *Cognition*, **21**, 37–46.

Baron-Cohen, S., Leslie, A.M., & Frith, U. (1986). Mechanical behavioral and intentional understanding of picture stories in autistic children. *British Journal of Developmental Psychology*, **4**, 113–125.

Barreau, S., & Morton, J. (1999). Pulling smarties out of a bag: a Headed Records analysis of children's recall of their own past beliefs. *Cognition*, **73**, 65–87.

Burgess, P. (2000). Theory and methodology in executive function research. In P. Rabbitt (Ed.), *Methodology of frontal and executive function* (pp. 81–116). Hove, UK: Psychology Press.

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

Carlson, M.S., Moses, L.J., & Hix, H.R. (1998). The role of inhibitory processes in young children's difficulties with deception and false belief. *Child Development*, **69**, 672–691.

Charman, T., & Baron-Cohen, S. (1995). Understanding photos, models and beliefs: a test of the modularity thesis theory of mind. *Cognitive Development*, **10**, 287–298.

Davis, H.L., & Pratt, C. (1995). The development of children's theory of mind: the working memory explanation. *Australian Journal of Psychology*, **47**, 25–31.

Freeman, N., & Lacohée, H. (1995). Making explicit 3-year-olds' implicit competence with their own false beliefs. *Cognition*, **56**, 31–60.

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

German, T.P., & Leslie, A.M. (2000). Attending to and reasoning about mental states. In P. Mitchell & K. Riggs (Eds.), *Children's reasoning and the mind* (pp. 229–252). Hove, UK: Psychology Press.

Gerstadt, C.L., Hong, Y.J., & Diamond, A. (1994). The relationship between cognition and action: performance of children $3^1/_2$–7 years old on a Stroop-like day-night test. *Cognition*, **53**, 129–153.

Gopnik, A., & Astington, J. (1988). Children's understanding of representational change and its relation to the understanding of false belief and appearance–reality distinction. *Child Development*, **59**, 26–37.

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.

Harris, P.L. (1992). From simulation to folk psychology: the case for development. *Mind & Language*, **7**, 120–144.

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

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

Harris, P.L., & Leevers, H.J. (2000). Reasoning from false premises. In P. Mitchell & K. Riggs (Eds.), *Children's reasoning and the mind* (pp. 67–86). Hove, UK: Psychology Press.

Harris, P.L., & Núnez, M. (1996). Understanding of permission rules by pre-school children. *Child Development*, **67**, 1572–1591.

Kahneman, D., & Tversky, A. (1982). The simulation heuristic. In D. Kahneman, P. Slovic & A. Tversky (Eds.), *Judgment under uncertainty* (pp. 201–208). Cambridge: Cambridge University Press.

Leslie, A.M. (1987). Pretense and representation: the origins of 'theory of mind'. *Psychological Review*, **94**, 412–426.

Leslie, A.M. (1994). Pretending and believing: issues in the theory of ToMM. *Cognition*, **50**, 211–238.

Leslie, A.M., German, T.P., & Pollizi, P. (in preparation). Belief-desire reasoning as a process of selection.

Leslie, A.M., & Pollizi, P. (1998). Inhibitory processing in the false belief task: two conjectures. *Developmental Science*, **1**, 247–253.

Leslie, A.M., & Thaiss, L. (1992). Domain specificity in conceptual development: neuropsychological evidence from autism. *Cognition*, **43**, 225–251.

Lezak, M.D. (1983). *Neuropsychological assessment* (2nd edn.). Oxford: Oxford University Press.

Luria, A.R. (1966). *Higher cortical functions in man*. London: Tavistock.

Mitchell, P., & Lacohée, H. (1991). Children's early understanding of false belief. *Cognition*, **39**, 107–127.

Moses, L. (2001). Executive accounts of theory of mind development. *Child Development*, **72**, 688–690.

Nichols, S., & Stich, S. (forthcoming). *Mindreading*. Oxford: Oxford University Press.

Perner, J., Leekham, S.R., & Wimmer, H. (1987). Three-year-olds' difficulty with false belief: the case for a conceptual deficit. *British Journal of Developmental Psychology*, **5**, 125–137.

Peterson, D.M., & Bowler, D.M. (2000). Counterfactual reasoning and false belief understanding in children with autism. *Autism*, **4**, 391–405.

Peterson, D.M., & Riggs, K.J. (1999). Adaptive modelling and mindreading. *Mind & Language*, **14**, 80–112.

Rabbitt, P. (2000). Introduction: methodologies and models in the study of executive function. In P. Rabbitt (Ed.), *Methodology of frontal and executive function* (pp. 1–38). Hove, UK: Psychology Press.

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

Riggs, K.J., Peterson, D.M., Robinson, E.J., & Mitchell, P. (1998). Are errors in false belief tasks symptomatic of a broader difficulty with counterfactuality? *Cognitive Development*, **13**, 73–91.

Robinson, E.J., & Beck, S. (2000). What is difficult about counterfactual reasoning? In P. Mitchell & K. Riggs (Eds.), *Children's reasoning and the mind* (pp. 101–119). Hove, UK: Psychology Press.

Russell, J., Saltmarsh, R., & Hill, E. (1999). What do executive factors contribute to the failure of false belief tasks by children with autism? *Journal of Child Psychiatry and Psychology*, **40**, 859–868.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: representation and the containing function of wrong beliefs in young children's understanding of deception. *Cognition*, **13**, 103–128.

Zaitchik, D. (1990). When representations conflict with reality: the pre-schooler's problem with false beliefs and 'false' photographs. *Cognition*, **35**, 41–68.

Received: 15 October 2001
Accepted: 2 July 2002