



## Competence and performance in belief-desire reasoning across two cultures: The truth, the whole truth and nothing but the truth about false belief?

Amir Amin Yazdi<sup>a</sup>, Tim P. German<sup>b,\*</sup>,  
Margaret Anne Defeyter<sup>c</sup>, Michael Siegal<sup>d,\*</sup>

<sup>a</sup>*Department of Psychology, Ferdowsi University, Mashad, Iran*

<sup>b</sup>*Department of Psychology, University of California, Santa Barbara, CA 93106-9660, USA*

<sup>c</sup>*Division of Psychology, University of Northumbria, Newcastle upon Tyne, UK*

<sup>d</sup>*Department of Psychology, University of Sheffield, Western Bank, Sheffield S10 2TP, UK*

Received 3 May 2004; accepted 23 May 2005

---

### Abstract

There is a change in false belief task performance across the 3–5 year age range, as confirmed in a recent meta-analysis [Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory mind development: The truth about false-belief. *Child Development*, 72, 655–684]. This meta-analysis identified several performance factors influencing success, including manipulations that highlight the salience of the initial belief content (such as asking where Sally will look *first* for the marble). However, because a proportion of variance in performance remained unexplained even when identified performance factors were controlled for, the authors concluded from the standpoint of a ‘theory–theory’ account that children’s improvement is the result of conceptual change. Further, the meta-analysis showed that manipulations such as ‘look first’ improve performance only in children who are in the older part of the 3–5 year range, and thus plausibly operating with a ‘transitional’ theory of mind—just on the point of realizing conceptual change. Here, we present three studies systematically investigating the ‘look first’ manipulation which showed that: (i) the advantage for the look first question can be demonstrated in children across different cultures, (ii) look first has an effect that is additive to the improvement with age; there is no interaction such that older children gain more benefit from younger children, (iii) performance in younger children can be, but is not

---

\* Corresponding authors. Fax: +1 805 893 4303 (TG); +44 144 2766515 (MS).

E-mail addresses: [german@psych.ucsb.edu](mailto:german@psych.ucsb.edu) (T.P. German), [m.siegal@sheffield.ac.uk](mailto:m.siegal@sheffield.ac.uk) (M. Siegal).

always, elevated to levels that are statistically above chance. These results challenge the theory–theory account and are discussed in terms of models of belief-desire reasoning in which *both* conceptual competence *and* performance factors play central roles.

© 2005 Elsevier B.V. All rights reserved.

*Keywords:* Belief-desire reasoning; Theory of mind; Competence-performance; Theory–theory; Look-first; Modularity

---

“The truth is never pure and rarely simple”—Oscar Wilde.

## 1. Introduction

A major achievement in the domain of mental state reasoning concerns the capacity to solve problems in which the action of a character with a *false* belief must be predicted. For example, in the ‘Sally-Anne’ task (Baron-Cohen, Leslie, & Frith, 1985), children are told about a story character (Sally) who has a false belief about the location of a marble. The character is described as having placed the marble in a box but, when she is away, another story character (Anne) moves it into a different location. The test question concerns where Sally will look for the marble. Research has typically shown that 4- and 5-year-old often succeed on such “standard” false belief tasks, whereas 3-year-old do not (see Wellman, Cross, & Watson, 2001; for a review). The underlying mechanism that permits a shift from incorrect to correct responses on false belief tasks between these ages is still controversial (Bloom & German, 2000; Moses, 2001; Scholl & Leslie, 2001).

Two main accounts aim to explain the improvement across this age range. According to the theory–theory account (e.g. Wellman et al., 2001), there is a conceptual change in children’s understanding of mind that underwrites the improvement; 3-year-old’s failure on false belief tasks can be attributed to a conceptual deficit in their ‘theory of mind’ (ToM). Older children (e.g. 4-year-old), most of whom answer correctly on false belief tasks, are assumed to have constructed a mature theoretical understanding of what beliefs really are, possibly via domain general mechanisms of theory formation (Gopnik, 1993, 1996; Gopnik & Astington, 1988; Gopnik & Wellman, 1992, 1994; Perner, 1991, 1995; Wellman, 1990; 2002; Wellman et al., 2001).

In contrast, an alternative account suggests that mental state reasoning is underwritten by a modular mechanism that is domain-specific and reliably developing, and which forms the basis of the capacity to first attend to and then learn about mental states (German & Leslie, 2000, 2001, Leslie, 1994a; Leslie, Friedman, & German, 2004; Scholl & Leslie, 1999, 2001). On this account, theory of mind concepts do not embody *theoretical* understanding of mental states, but rather serve to token instances of the relevant mental state in the world; they allow children to attend to the properties of believing, desiring and pretending, and thus learn about them (see German & Leslie, 2000, 2001; Leslie, 2000; Leslie et al., 2004; Leslie, German, & Pollizi, 2005, for extended discussion of this idea; see Fodor, 1998, for a more general treatment of concepts that is consistent with this view).

On this latter story, mental state concepts are early and reliably expressed, coming on line in typically developing children, at the latest, sometime within the second year as indexed by the advent of pretend play (Leslie, 1987, 1994b, 2000). Thus something else must account for the shift in performance in the ability to pass false belief tasks at about age four. Leslie and his colleagues have proposed processing models of successful false belief reasoning in which gradual increase in executive selection processing capacities allows children to overcome features of belief desire reasoning problems that cause consistent errors (e.g. the default tendency to attribute a *true* belief in the false belief task; Bloom & German, 2000).

There are increasing numbers of findings demonstrating that false belief performance in young children is improved by manipulations that reduce executive selection demands (Birch & Bloom, 2003; Freeman & Laco  e, 1995; Happ   & Loth, 2002; Mitchell & Laco  e, 1991; Roth & Leslie, 1991; Siegal & Beattie, 1991; Surian & Leslie, 1999; Zaitchik, 1991), as well as more recent demonstrations that increases in the potency of the same kinds of factors can cause problems for older children otherwise capable of solving ‘standard’ false belief tasks (Cassidy, 1998; Friedman & Leslie, 2004a, 2004b, 2005; Leslie et al., 2005; Leslie & Pollizi, 1998). Even adults’ capacities to reason effectively about the knowledge states of others can be stressed in cases where they must attribute lack of knowledge in situations where they are knowledgeable (e.g. Bernstein, Atance, Loftus, & Meltzoff, 2004; Birch & Bloom, 2004; Fischhoff, 1975).

There are also many findings linking false belief performance with a number of more general ‘executive’ abilities, including inhibitory processing (Carlson & Moses, 2001; Carlson, Moses, & Breton, 2002; Carlson, Moses, & Hix, 1998). These findings have engendered different specific performance accounts, but all share the basic notion that there will be manipulations of the standard false belief task that will have profound affects on performance, suggesting that conceptual change may play little role in explaining this particular developmental shift in the domain of theory of mind.<sup>1</sup>

Note that ‘theory–theory’ accounts of mental state reasoning and the development thereof also acknowledge a role for performance demands. However, under the conceptual change view such factors play only a peripheral role. As Wellman et al. put it:

Performance on a cognitive task reflects at least two factors: conceptual understanding required to solve the problem (“competence”) and *other non-focal cognitive skills* (remembering key information, focus attention, comprehend, answer questions) required to access and express understanding (performance).  
” (Wellman et al., 2001, pp. 656–657, italics added).

The meta-analysis conducted by Wellman et al. (2001) documents the importance of performance factors in the solution of false belief tasks. However, because there was no single performance factor that removed the difficulty for the younger child, Wellman et al.

---

<sup>1</sup> A full and extended discussion of the various performance accounts is beyond the scope of the current paper. The interested reader is referred to the excellent discussion by Moses and Carlson (2001); see also Carlson et al. (2002) introducing the distinction between ‘expression’ and ‘emergence’ accounts of the relationship between executive function accounts and ‘theory of mind’. For a different take on this same debate, see Perner & Lang (2000). For a discussion of the relationship among difference performance accounts, see Leslie et al. (2005).

concluded that the ‘truth’ about false belief is that conceptual change must play a role in explaining the shift between ages three and four. Moreover, Wellman and colleagues also claims that some putative performance manipulations are effective *only* for children at the older end of the 3–4-year range, and thus already close to achieving conceptual change—i.e. children who are in the ‘transitional’ stages of re-organizing their conceptual system in the appropriate way (Gopnik, 1996; Gopnik & Wellman, 1992, 1994). In the current investigation, we address these two specific conclusions by focusing on the role of one particular performance manipulation—Siegal and Beattie’s (1991) ‘look first’ procedure—on children across the entire 3–4 years of age range.

To foreshadow our conclusions, across three studies we show that the conclusions drawn from the Wellman et al. meta-analysis, as a statement on the truth about false belief, though widely accepted as a definitive statement on the theoretical debates in this domain, do not stand up. The ‘look first’ procedure turns out to be effective at all points of the 3–4 year range, to be effective across both Western industrial and non-western developing cultures, and when tested via different specific false belief tasks and when assessed both between and within subjects. The lack of interaction observed here between kind of questioning (‘standard’ versus ‘look first’) and age is inconsistent with the predictions made by theory–theory accounts of the acquisition of belief-desire reasoning.

### 1.1. The ‘look first’ manipulation

One of the first demonstrations of a performance manipulation with a profound effect on children’s success in solving false belief problems was reported by Siegal and Beattie (1991). They designed two experiments to investigate responses on false belief tasks in terms of children’s awareness of the purpose and relevance of test questions about the beliefs of a story character. They hypothesized that when an experimenter in the experimental setting asks questions such as “When will Sally look for the object?” 3-year-olds may not share the experimenter’s intention that the question refers to how a person with a false belief will initially be misled. Instead, children may assume that the question refers to where Sally must or should look to find the object.

For this reason, Siegal and Beattie (1991) changed the question in the false-belief task from “Where will (the character with the false belief) look for the object?” to “Where will the character look *first* to find the object?” They argued that, as a consequence, young children are more likely to share the purpose and relevance of experimenter’s question. Their results showed that by adding the qualifier “first”, 3-year-olds’ performance improved significantly to a level that was above chance. This result was initially characterized as stemming from a ‘clarification’ in the experimenter’s intended question; the look first version of the task lowers the likelihood that the 3-year-old will interpret the test question not as “Where will the character look (first) for the object?” as intended by the experimenter but instead will gloss it as something like: “Where should the character go to find the object?” (Siegal & Beattie, 1991).

The ‘look first’ advantage has been replicated, and also recently characterized as achieving its facilitating role in terms of reducing the executive inhibitory demands present in the standard version of the task (Leslie et al., 2005; Surian & Leslie, 1999). To make this clear, we adopt the theoretical framework of Leslie and colleagues (German &

Leslie, 2000, 2001, 2004; Leslie, 1994a, 2000; Leslie & Thaiss, 1992; Leslie et al., 2005, German, & Pollizi, in press).

Under this view, the early developing competence is supplied by a modular mechanism (the Theory of Mind Mechanism, or ToMM) that supplies basic mental state concepts (e.g. *belief*, *desire*, *pretend*) that allow the child to attend to behavior and infer the underlying mental state that gave rise to it. ToMM supplies a small number of candidate belief state contents for a given stream of behavior, and always includes a true belief content as a default. The true belief content is always supplied because in the absence of any other information, the only constraint on the protagonist's belief is the current true state of affairs (from the point of view of the attributer).

In most cases, the default attribution will not pose a problem because most of the time people's beliefs *are* true. From the perspective of conversational understanding between speakers and listeners, speakers normally abide by maxims or rules such as "Tell the truth and avoid falsehood" that prompt listeners to assume the truth of beliefs (Grice, 1975). But in some instances, speakers depart from such rules and so induce false beliefs in others. In such unusual situations where a false belief has been acquired, the default attribution must be *inhibited* and an alternative non-factual content for the belief selected instead.

These performance demands are assumed to be handled by another mechanism, the Selection Processor (SP), which is a kind of executive function more domain general in character (at least, penetrable by more information than is ToMM). SP may not necessarily be specific to the domain of belief desire reasoning (see, e.g. Leslie & Thaiss, 1992; Roth & Leslie, 1998), though it is also *possible* that the inhibition and selection process is to some extent domain-specific (see Leslie et al., 2005).

Using this analysis, Surian and Leslie (1999) explain how the "look first" question improves children's performance on false belief tasks. They argue that, in belief attribution, belief contents that are true are attributed *by default*. In a false belief task, the default true-contents leads to a wrong answer and therefore needs to be *inhibited*. By increasing the salience of the first location as the possible content of Sally's belief, the 'look first' question reduces the need for inhibition. Therefore the "look first" false belief task places less of a load on the SP and enhances children's performance in comparison with their performance on standard false belief tasks. In this analysis, children as young as 3 years are able to represent beliefs. Their difficulty on standard false belief task results from limitations in access to selection resources that gradually develop over the preschool period and beyond.

### 1.2. *Look first and meta-analysis*

Wellman et al.'s meta-analysis deals only with findings on look first (and similar manipulations aimed at increasing salience of the false belief content in other false belief tasks) published up to January 1998 (e.g. Freeman & Laco  e, 1995; Lewis & Osborne, 1990; Siegal & Beattie, 1991). The analysis excludes more recent results showing that explicit questioning on false belief tasks does enhance 3-year-olds' performance (Joseph, 1998; Nelson et al., 2003; Surian & Leslie, 1999).

Wellman et al., as noted above, maintain that "temporal marking" manipulations, as they term changes such as the look first procedure, do not change the basic developmental

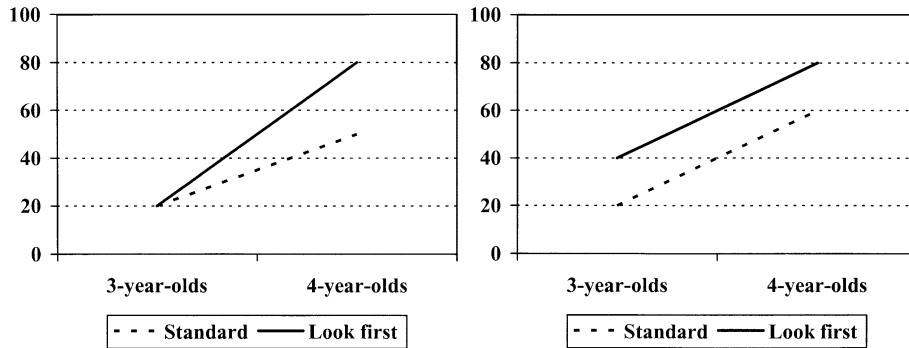


Fig. 1. Hypothesized pattern of performance on 'standard' and 'look first' versions of the false belief task according to 'theory-theory' (see Wellman et al., 2001), in which there is an age  $\times$  condition interaction (left panel). Right panel shows predicted pattern of performance according to the ToMM-SP model (see, e.g. Scholl & Leslie, 2001).

pattern of false belief responses in preschoolers. In terms of the theory-theory account, performance manipulations only facilitate ToM performance once children at the older end of the 3- to 4-year-old spectrum are in the pre-conceptual change transitional period—and thus have essentially the conceptual resources to answer false belief tasks correctly. This idea has been articulated in terms of children having discovered the concept of belief as an 'auxiliary hypothesis' to their initial (and wrong) theory (Gopnik & Wellman, 1994). According to this view, only among older 3-year-olds and young 4-year-olds, should look first questioning significantly increase correct responses on the false belief task to a level above what would be expected by chance. Thus, there should be an age  $\times$  questioning format interaction effect in children's responses to standard and look first conditions of false belief tasks (see Fig. 1, left panel).

In contrast to the theory-theory position, the hypothesized ToMM/SP pattern of children's responses on standard and look first false belief task suggests that there should be a relatively stable facilitating effect on false belief performance across the 3–4 year range. According to Scholl and Leslie (2001), the look first manipulation can be thought of as adding a 'constant' amount of help to the inhibitory problem, which should increase the probability of success on a given trial by some measurable amount. This is depicted in the right panel of Fig. 1, which shows main effects for age and condition ('standard' versus look first) and no interaction. According to this account, reduction of performance demands through look first questioning leads to increased success on false belief tasks at all ages across the relevant age range.

In the first investigation reported here, we sought to adjudicate between these alternatives using a large sample of children that would permit a more detailed analysis of response patterns. In follow-up studies, we addressed two further questions: (i) whether the advantage for look first questions over standard questions was the result of false positives—success achieved via general biases toward responding to the first location, irrespective of what was the character's belief, and (ii) whether the effect of look first can be demonstrated using a within subjects experimental design, and specifically in a group of

children who are at the younger end of the 3–4-year-old range (and thus not plausibly in the ‘transitional’ stage of conceptual development prior to theory change).

At the same time, since previous studies on look first effects have been limited to industrialized Western cultures, our aim was to determine whether look first effects emerge in children living in the culture of a developing country. In our first experiment, we aimed to re-examine the Western pattern of children’s responses on standard and “look first” false belief tasks by attempting to replicate Siegal and Beattie’s (1991) findings in Iran.

## 2. Experiment 1

### 2.1. Method

#### 2.1.1. Participants

These were 194 children divided into two age groups of 90 3-year-olds (mean age 3 year, 5 months, range 36–47) and 104 4 year-olds (mean age 4 year, 4 months, range 49–60). The children attended preschools located in middle class areas of Mashad, Iran. All children had parental permission to participate. In addition, nine children, two 3-year-olds and seven 4-year-olds, answered incorrectly on reality control questions. The data from these children were discarded from final analyses. There were approximately equal numbers of boys and girls assigned to each condition.

#### 2.1.2. Materials and procedure

As in Siegal and Beattie’s (1991) Experiment 1, in each age group, the children were randomly assigned to one of two conditions: standard or look first. Each child was tested individually on two stories. In order to make use of false belief tasks with Iranian children, the stories were initially translated into the Persian language and then the Persian version was translated back into English to ensure that the Persian version corresponded with the English one.

For the children in the standard condition, the stories were similar to those used in standard false belief tasks (Wellman & Bartsch, 1988): (1) “Maryam wants to find her kitten. Maryam thinks her kitten is in the kitchen. Maryam’s kitten is really in the bathroom. Where will Maryam look for her kitten? Where is it really?” (2) “Ali wants to find his puppy. Ali’s puppy is really in the kitchen. Ali thinks his puppy is in the bathroom. Where will Ali look for his puppy? Where is it really?” The locations of the pets in the stories were counterbalanced across subjects so that, for half the children, Maryam was said to believe that her kitten was in the kitchen (real location: bathroom) and Ali to believe that his puppy was in the bathroom (real location: kitchen); for the other half, the real and believed locations were reversed. The procedure for the look first condition was identical to that of the standard condition except that the children were asked, “Where will Maryam (Ali) look first for her (his) kitten (puppy)?”

Within each age group, the order of the stories was counterbalanced, as were the orders of the premises concerning beliefs and reality. During the story presentation, the experimenter referred to miniature figures of a girl, boy, kitten, and puppy, and to locations inside a large wooden doll house (60 × 60 cm).

## 2.2. Results

Children's responses on false belief tasks were scored on a 0–2 scale with each correct answer receiving one point. The scores were analyzed in a 2 (Age) × 2 (Condition) factorial ANOVA. There were significant main effects for age,  $F(1,194) = 6.52$ ,  $P < 0.01$  and condition,  $F(1,194) = 0.65$ ,  $P < 0.01$ . The age × condition interaction effect was not significant,  $F(1,194) = 0.65$ ,  $P = 0.41$ . The 4-year-olds outperformed 3-year-olds with mean of 1.31 out of maximum possible score of 2 compared to 3-year-olds' mean score of .65. Children questioned in the look first condition outperformed those in the standard condition. The mean score in the look first condition was 1.32 out of a maximum possible of two correct compared to the mean in the standard condition of .68.

Consistency in the children's responses on false belief tasks is shown in Table 1. Children in the standard condition were consistent in answering incorrectly to false belief tasks. Of the 45 3-year-olds in the standard condition, 4 were correct on both stories, 5 on one story, and 36 on neither, a performance level that is far below chance ( $P < .01$ , binomial test). In contrast, those in the look first condition performed at an above chance level. Out of 45 3-year-olds, 17 were correct on both stories, 12 were correct on one story, and 16 on neither. Assuming a 25% probability of solving both problems successfully by chance, the numbers of children answering correctly on both stories exceeded chance,  $z = 1.807$ ,  $P = 0.0307$ . The 4-year-olds in both conditions were consistent in answering correctly. In the look first condition, out of 53 4-year-olds, 40 were correct in both stories, 4 were correct on one story, and 9 on neither. Comparable numbers for the standard condition were 22, 9, and 20 (Fig. 2).

Altogether the children's responses conformed to the pattern predicted by the ToMM/SP position (see Fig. 1). Even at the youngest ages, there was a significant effect for condition (Fig. 3). Of the 22 children aged 36–41 months in the standard condition, 20 children answered incorrectly on both tasks; one child was correct on both tasks and another was correct on one of the two tasks. In contrast, of the 25 children aged 36–41 months in the look first condition, only 11 children answered incorrectly on both tasks; 6 children were correct on both tasks and 8 were correct on one task.

## 2.3. Discussion

Consistent with the ToMM/SP hypothesis, the results support the notion that the look first improves both 3- and 4-year-olds' performance on false belief tasks. This is in

Table 1

Percentage of children scoring both correct, one correct and none correct on the two false belief tasks in the standard and look first questioning conditions of Experiment 1 (Siegal and Beattie's (1991) results are shown in parentheses)

	Standard		Look first	
	3-year-olds, <i>N</i> = 45 (20)	4-year-olds, <i>N</i> = 51 (20)	3-year-olds, <i>N</i> = 45 (20)	4-year-olds, <i>N</i> = 53 (20)
Both correct	09 (20)	43 (30)	37 (60)	75 (65)
One correct	11 (20)	18 (20)	27 (25)	08 (10)
Both incorrect	80 (60)	39 (50)	36 (15)	17 (25)



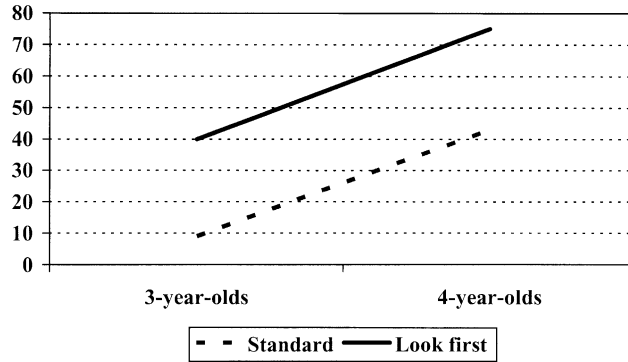


Fig. 2. Percentage of children correct in Experiment 1 on both false belief trials, according to age and question type ('standard' versus 'look first').

contrast to the theory–theory account in which 3-year-olds’ performance is supposed not to be improved in the look first condition.

Concerning the role of “look first” on children’s performance on false belief tasks, the findings are consistent with the results obtained from Siegal and Beattie’s (1991) Experiment 1 in which children gave 71% correct answers in the look first condition and only 30% correct answers in the standard condition. The comparable results for the current experiment were 57% correct answers in the look first and 32% in the standard condition. In both Siegal and Beattie’s Experiment 1 and the current experiment, children performed better in the look first than the standard false belief tasks. In both Siegal and Beattie’s Experiment 1 and the current experiment, the 3-year-olds’ performance on standard false belief tasks was far below what is expected by chance. While 3-year-olds in the look first condition who were in Siegal and Beattie’s (1991) Experiment 1 performed at an above chance level, the performance of 3-year-olds in the current experiment was no greater than chance.

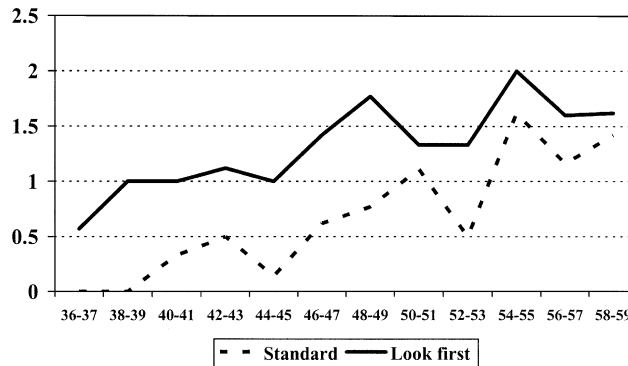


Fig. 3. Performance of children in Experiment 1 in ‘standard’ and ‘look first’ conditions according to age (divided into 2-month bins).

Unlike the results of Siegal and Beattie (1991), the results of the current experiment also indicate the main effect of age demonstrating that 4-year-olds outperformed 3-year-olds' performance on false belief tasks. Iranian 3-year-olds' performance on both the standard and look first conditions in this experiment was slightly worse than that of the Western children tested by Siegal and Beattie. The difference may be due to the fact that the Iranian children (mean age 3 year, 3 months) were on an average younger than Western children (age mean 3 year, 8 months) in Siegal and Beattie's (1991) Experiment 1.

However, it might be argued that the experimenter's use of the term "first" might have somehow signaled to children that they should give answers that are contrary to the real location of the desired object. If so, the children then could succeed on the task without attending to the character's beliefs. The purpose of Experiment 2 was to replicate the results of Siegal and Beattie's (1991) Experiment 2 in which a true belief control task was used in order to guard against the possibility of false positive responses.

On the one hand, if children interpret the look first question as generically implying that they should give an answer different from the real location of the object, then they should respond correctly in the look first condition of false belief tasks but not in the look first condition of true belief tasks. On the other hand, if they give correct answers on both look first false and true belief tasks as in Siegal and Beattie's (1991) Experiment 2, they would be interpreting the term first in a manner sensitive to the character's belief state, ruling out the false positive interpretation. By aiming to replicate the results of Siegal and Beattie's (1991) Experiment 2 (see also Surian & Leslie, 1999), we sought to investigate whether "first" creates its effect by inducing false positive responses or whether it truly leads to better belief calculation.

### 3. Experiment 2

#### 3.1. Method

##### 3.1.1. Participants

The subjects were 22 3-year-olds with a mean age of 3 year, 6 months, range 3:1–3:11. An additional 11 children, 7 in the "look first" condition and 4 in the standard condition, answered reality control questions incorrectly and were discarded from final analyses. The children attended preschools located in middle class areas of Mashad, Iran. There were approximately equal numbers of boys and girls in each condition. All had parental permission to participate. No child had participated in Experiment 1

##### 3.1.2. Material and procedures

As in Experiment 1, the children were randomly assigned to one of two groups: the standard condition and the look first condition. For the standard condition group ( $n=13$ ), the tasks were to predict the actions of a story character based on a true belief (TB) or a false belief (FB). For the look first condition group ( $n=9$ ), the tasks were the same but involved predicting where the story character would look first either in regard to TB or FB.

The TB story was "Maryam wants to find her kitten. The kitten lives in two rooms: the garage and the lounge. Maryam thinks her kitten is in the garage and now it really is in the

garage. Where will Maryam look for her kitten? (test question) Where is the other room that the kitten lives in?” (control question). The false belief story was similar to that used in Experiment 1: “Ali wants to find his puppy. Ali’s puppy is really in the kitchen. Ali thinks his puppy is in the bathroom. Where will Ali look for his puppy? Where is it really?” The content of the TB and FB stories in the standard condition was identical to TB and FB stories in the look first condition, respectively, except that the children in the “look first” condition were asked, “Where will Maryam (Ali) look first for her kitten (his puppy)?”

The locations of the pets in the stories were counterbalanced across subjects so that, for half the children, Maryam was said to believe that her kitten was in the lounge (real location: lounge) and Ali to believe that his puppy was in the kitchen (real location: bathroom); for the other half, the real and believed locations were reversed. The orders of the premises within the stories were also counterbalanced. The same materials as in Experiment 1 accompanied the stories.

### 3.2. Results

Table 2 shows children’s responses to test questions in the standard and look first conditions. Children in the look first condition outperformed children in the standard one. While seven of the nine children in the look first condition recognized that the character would look in the wrong location when the belief was false, the comparable number in the standard condition was only 2 of 13. The difference between the groups was significant ( $P < 0.05$ , Fisher exact probability test). In contrast, there was no significant difference between children’s performance on standard and look first TB tasks. In this condition, most of children responded that the character that had a true belief would look where he or she thought the desired pet was located. Of the 13 and 9 children in the standard and look first conditions, respectively, the numbers correct on the TB tasks were 13 and 7, respectively.

The children’s performance on FB and TB tasks in look first condition, however, was similar. Of the nine children in the look first condition, the numbers correct on the TB and FB tasks, respectively, were 7 and 7. The similarity might be explained by the fact that children on the FB task performed at ceiling.

### 3.3. Discussion

The results support the hypothesis that children’s better performance on the look first condition does not stem from inducing false positives—correct responses that are caused

Table 2

Responses of the children on the false belief and true belief tasks in the standard and look first questioning conditions of Experiment 2 (Siegal and Beattie’s (1991) results in parentheses)

	Standard		Look First	
	FB, $N = 13$ (12)	TB, $N = 13$ (12)	FB, $N = 9$ (12)	TB, $N = 9$ (12)
Correct	15 (42)	100 (92)	78 (83)	78 (83)
Incorrect	85 (58)	0 (08)	22 (17)	22 (17)

by a bias to select the first location irrespective of the belief state of the character. If a general ‘first location bias’ had been introduced by the look first procedure, it should be applied whether it is the right answer (as in the case of the false belief stories) or the wrong answer (as in the true belief control stories). The results of Experiment 2 replicate the results of both Siegal and Beattie’s (1991) Experiment 1 and Surian and Leslie (1999), as well as those of Experiment 1 in showing that children often respond correctly when asked to predict the initial behavior of a story character with a false belief. These results do not support the alternative hypothesis that children misinterpret the word *first* and perceive the question “Where will she look for first?” to require an answer that is other than the real location of the desired object. If a general bias to respond toward the empty location was introduced, children ought to have made errors in the true belief control condition with the look first question—a result that did not occur in this experiment.

The present Experiments 1 and 2, taken together, show that children appear to be helped in attending to the first location as a possible target content for the character’s belief. This result was initially characterized (e.g. in Siegal & Beattie, 1991) as stemming from a ‘clarification’ to the experimenter’s intended question; the look first version of the task lowers the likelihood that the 3-year-old will interpret the test question not as “Where will the character look (first) for the object?” as intended by the experimenter but instead will gloss it as something like: “Where should the character go to find the object?” (Siegal & Beattie, 1991).

As argued in Section 1, within the ToMM-SP framework, we can further articulate the mechanism via which this clarification might work. Namely, we propose that the look first procedure increases the salience of the false belief content such that less inhibition is required to deselect the ‘true belief’ content and attribute the false belief content. Given the limited resources assumed to be available to 3-year-olds, this help results in more children solving the belief problem in this case, than are able to do so when no such additional salience is attached to the false belief content (see also Leslie et al., 2005; Scholl & Leslie, 1999; Surian & Leslie, 1999).

In Experiment 3, we attempted to provide further evidence for the effect of the look first manipulation, and its role across the 3–4-year-old age range. In this experiment, we were particularly concerned with showing that the effect of look first is stable and appreciable *even among the youngest 3-year-old children*. Thus, we tested a sample of young 3-year-olds only, and compared them to a group of young 4-year-olds. This experiment thus provides a focused test of Wellman et al.’s (2001) claim that the look first manipulation works only for children who are in the transitional period immediately prior to conceptual change.

We tested the effect of the look first manipulation with a couple of changes to the materials and procedure of the false belief task, and in a group of children from a western cultural setting. The first change was to follow Surian and Leslie (1999) and test the look first procedure in a task modeled on the standard ‘change of location’ task (Baron-Cohen et al., 1985), rather than with tasks modified from Wellman and Bartsch (1988). We chose a picture book version of the Sally–Anne task for this purpose.

The second change to the procedure in this experiment was that we dealt with the issue of false positives by using a true belief task (as reported in Experiment 2). However, we used this task as a screening task, such that only children who solved both ‘think’ and ‘look

first' versions of the true belief task were admitted to the false belief study. In this way, we were able to be sure that none of our subjects were susceptible to general biasing effects of the look first question.

The final change was to assess the effect of look first within the same subjects. Because of possible interference caused by asking both 'look' and look first questions to the same children in quick succession, we achieved the within subjects procedure by using a think question as our standard assessment of false belief, and comparing this to look first.<sup>2</sup> We speculated that in a within subjects design, any advantage for look first might have some carry over effect for the think question, if children are induced by look first into being more likely to calculate a false belief content, and on the assumption that this information is available to a subsequent belief calculation process (one might conceive this as a kind of priming). Similarly, receiving the think question first (and calculating it incorrectly) might have a lingering effect on the activation of the true belief content, making it more highly activated, and reducing the effect of look first manipulation on easing the subsequent action prediction.

## 4. Experiment 3

### 4.1. Method

#### 4.1.1. Participants

The participants were 70 children aged between 3 years 0 months and 4 years 3 months. Of these children, 18 failed to solve the TB screening task, either by failing the look first question (3 children), the think question (11 children) or both (4 children), resulting in a final sample of 52, divided into a 'young three' year old group ( $N=22$ , mean age 3–3, range 3–0 to 3–5) and a 'young four' year old group ( $N=30$ , mean age 4–3, range 3–11 to 4–7).

There were approximately equal numbers of boys and girls in each age group, and the children's ethnicities reflected a typical urban northern city in the UK. All children had English as a first language.

#### 4.1.2. Materials and procedure

Two picture book tasks were produced, each with seven pages, conveying the main events of each task (true belief and false belief). The pictures, stories, questions and questions used in Experiment 3 are presented in Fig. 4. Children were randomly assigned to receive either the think question first or the look first question first for both the true belief screening and false belief task.

Children were tested individually, seated at a low table in a quiet area of their preschool. The children were first presented with the true belief-screening task. The story

---

<sup>2</sup> While there are no appreciable differences between the 'think' and 'look' versions of the false belief task (Wellman et al., 2001), it is important to note that this holds only for false belief tasks that also involve a desire to *approach* the object. A recent discovery is that action prediction is considerably harder than 'think' when the character's desire is to *avoid* the object in question (e.g. in the case where Sally wants to avoid the location of the marble, and has a false belief about its location; Cassidy, 1998; Friedman & Leslie, 2004a, 2004b; Leslie et al., 2005; Leslie & Pollizi, 1998).



One day, Bobby is in the kitchen eating a chocolate bar.



Then his mom comes into the kitchen and says, "Bobby, put away your chocolate bar. It's time for chores."



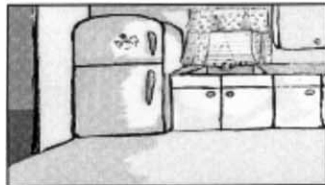
Bobby puts his chocolate bar inside the refrigerator and...



...asks his mom what he should do next. But while he is talking to her...



...his mom takes the chocolate from the refrigerator and puts it into the candy drawer.



MQ: Where did Bobby put the chocolate in the beginning?

RQ: Where is the chocolate now?



Later on, Bobby comes back into the kitchen and wants to eat his chocolate.

T: Where does Bobby think the chocolate is?

LF: Where will Bobby look first for the chocolate?

Fig. 4. Pictures, story and questions for True belief (panel A) and False belief (panel B) tasks used in Experiment 3.



One day, Sally is playing in her room with a ball.



Then her dad comes into her room and says, "Sally, put away your ball. It's time for breakfast."



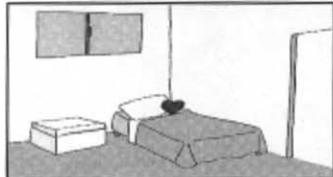
Sally puts her ball under her bed and...



...goes down for breakfast. But while she is away...



...her dad takes the ball from under the bed and puts it into the toy chest.



MQ: Where did Sally put the ball in the beginning?

RQ: Where is the ball now?



Later on, Sally comes back into her room and wants to play with her ball.

T: Where does Sally think the ball is?

LF: Where will Sally look first for her ball?

Fig. 4 (continued)

was presented one picture at a time, with the text (see Fig. 4) accompanying each story read out at that time. If children failed a control question (memory or reality), they were corrected, recycled through the story and asked again. If they failed a second time, this would have resulted in them being eliminated from the story, but in fact no child failed a second time. Children who solved the true belief task (defined as correctly answering both think and look first questions correctly) were included in the main study. After a short break in which the children were allowed to return to play in the class, those eligible for the false belief task were recalled and given the false belief story.

#### 4.2. Results

Children were scored correct if they answered that Sally would look for her marble under the bed (look first question) or thought that the marble was under the bed (think question). The number of children (and percentage of children passing the look first and think versions of each task) is shown in Table 3. There was a significant advantage of age for performance on the think question ( $X^2_{\text{corr}(1)} = 8.19$ ,  $P < 0.002$ , one-tailed), but this difference was not significant for the look first question ( $X^2_{\text{corr}(1)} < 2.00$ , NS).

The relation between performance on the think question and performance on the look first question is shown in Table 4. This shows that there were more children who passed the look first question while failing the think question than who showed the opposite pattern, a pattern that was significant for both 3-year-olds (McNemar binomial,  $N = 10$ ,  $k = 0$ ,  $P = 0.002$ , one-tailed) and for 4-year-olds (McNemar binomial,  $N = 5$ ,  $k = 0$ ,  $P = 0.036$ , one-tailed).

Finally, we investigated whether the advantage for look first was in any way affected by an order in which the questions came. This analysis revealed that while there was

Table 3

Percentage of children correct on ‘think’ and ‘look first’ versions of the false belief tasks in Experiment 3 (Surian and Leslie’s (1999) results in parentheses)

	Think	Look first
Young 3-year-olds	$N = 22$ (40) 14 (33)	$N = 22$ (40) 55 (58)
Young 4-year-olds	$N = 30$ (20) 57 (30)	$N = 30$ (20) 73 (83)

Table 4

Contingency between answers to ‘think’ and ‘look first’ versions of the false belief task in Experiment 3

		Look first question			
		Young 3-year-olds		Young 4-year-olds	
		Pass	Fail	Pass	Fail
Think question	Pass	3	0*	17	0*
	Fail	10*	9	5*	8

McNemar binomial  $P < 0.005$ , one-tailed;  $P < 0.05$ , one-tailed.



a tendency for children to solve the think and look first questions slightly better when the look first question was asked first (think question 26% correct when asked first, 48% when asked second; look first question 72% when asked first, 57% when asked second), these differences were not significant for either ‘think’ or look first questions ( $X^2_{\text{corr}(1)} < 2.00$ , NS, in each case).

#### 4.3. Discussion

The results of this experiment were clear. There was an advantage for look first over think versions of the false belief task that was demonstrated (i) within the same subjects, and (ii) within a group of subjects at the lower end of the 3–4 year age range. None of the 3-year-old children in this experiment were drawn from an age range close to the hypothesized ‘transitional’ periods prior to conceptual change (see e.g. Gopnik & Wellman, 1992, 1994). Nevertheless, there was an advantage for look first over think in this youngest group, as well as for the young 4-year-old group, who overall were far from ceiling on the standard think question (57%; improving to 73% on look first). Though this was a significant improvement via a one-tailed test, it was not as dramatic as that observed in the younger group, who improved from 14% performance on the think question to 55% on ‘look first’.

Thus, though effects of look first were demonstrated between both 3 and 4-year-olds in Experiment 3, this experiment came closest to showing an interaction between age and the effect of this performance factor. However, it is not the interaction that is predicted on Wellman et al.’s (2001) argument about the effect of look first, but rather hints at an interaction in which the *youngest* children gain more than their older peers, rather than *vice versa*.

### 5. General discussion

The results of the present research support the pattern of performance on look first questions hypothesized by the ToMM/SP theory, as depicted in Fig. 1. Recall that according to the ToMM/SP account, 3-year-olds’ difficulty in standard false belief tasks results from a performance limitation in overcoming a bias to attribute the default true belief content in false belief situations. If 3-year-olds’ performance is explained in this way, then the need to propose that the success of 4- and 5-year-olds is due to a conceptual shift success evaporates, and the main tenet of theory–theory account is overruled.

In our view, the early developing mechanism (ToMM) comes online during the second year of life and supplies basic attitude concepts (e.g. belief, desire, pretend) that allow the child to attend to behavior and interpret it as emanating from mental states. ToMM is limited in its ability to access information from all parts of the cognitive system, and effective belief-desire reasoning thus relies on the functioning of a further executive selection process to select among the range of candidate mental state contents ToMM calculates from perceptual descriptions of behavior.

The selection processor increases in its efficiency across development, producing a change in performance on standard tests of false belief in the preschool period and beyond.

Versions of the false belief task that introduce reduced selection demands (such as ‘look first’) create situations where children are probabilistically more likely to generate the correct response (Siegal & Beattie, 1991; Surian & Leslie, 1999). Versions of the false belief task that increase executive selection demands (such as combining the false belief with a desire to avoid the object) create situations where children are probabilistically less likely to generate the correct response (Cassidy, 1998; Friedman & Leslie, 2004a, 2004b; Leslie et al., 2005; Leslie & Pollizi, 1998). Interestingly, one recent demonstration shows that the facilitating effect of look first on the performance of 3-year-olds on standard false belief problems also extends to reducing problems for 4-year-olds on the false belief tasks that have increased inhibitory demands (Leslie et al., 2005).

The current studies provide more evidence on the specific nature and profile of the effect that the look first manipulation has on performance on false belief reasoning across the 3–4 year period, and in a culture as yet unstudied by researchers interested in belief-desire reasoning. Specifically, it shows that the effect of look first is *not* restricted to those children who are ‘closest’ to achieving a conceptual change (e.g. only older 3-year-olds or young 4-year-olds). Across all three studies, there was an appreciable advantage of look first over the standard form of question that extended to the youngest children in the sample. Moreover, in one study (Experiment 2) the advantage for younger children exceeded ‘chance’ (see below) while in another the advantage was, if anything, greater for young 3-year-olds than it was for young 4-year-olds (Experiment 3).

This specific pattern of results is not consistent with the conceptual change explanation of ‘look first’ and similar manipulations offered by Wellman et al. (2001) in discussing their meta-analysis from a largely theory–theory perspective. The pattern that Wellman et al. show in that analysis (where look first and similar manipulations help only older children) is likely skewed by the fact that the 1998 cut-off date excluded several look first type manipulations that focus particularly on 3-year-olds’ performance on false belief tasks (Joseph, 1998; Nelson et al., 2003; Surian & Leslie, 1999; see Scholl & Leslie, 2001).

Our results address several issues that are prominent in debate surrounding competence and performance in belief desire reasoning: first, the notion that improved performance on false belief is meaningful only if performance is elevated above ‘chance’, and second, how the look first manipulation can illuminate the question of how apparently different performance factors can be captured in a common theoretical framework.

### *5.1. Competence, performance and chance*

Considerable controversy in the ‘theory of mind’ literature has surrounded the proposal that demonstrating that a task manipulation leads to some improvement in false belief performance is itself not important unless the improvement in performance is to a point that is above chance performance (e.g. Wellman et al., 2001), where chance is assumed to be 50%, on the basis of the fact that there are typically two responses in the false belief task.

While we agree that demonstrating that performance statistically exceeds that expected if children were guessing suggests that the improved performance is not attributable to the manipulation causing children to respond at random; it is less clear that performance that

does not exceed chance is always best explained by assuming that children are ‘just guessing’.

It is important to note that while performance that exceeds ‘chance’ suggests that children are not responding randomly *with respect to the response options available*, it does not in itself show that the performance has improved for non-trivial reasons. For example, if a set of results shows that alongside perfect responding on a false belief problem with a new manipulation, the same children *all fail* a true belief problem with the same manipulation then this pattern of results suggests that the improvement is likely caused by some non-interesting bias to respond to one location. This is true, despite the fact that the performance in both conditions is significantly different from 50 to 50 or ‘chance’.

We note that 3-year-olds’ performance in our investigation did not rise to ‘above chance’ levels in Experiments 1 and 3 (though it was in Experiment 2). However, it seems that there are a number of good reasons to reject the idea that any responding that does not exceed 50% (where there are two options) is best explained by assuming that the children were guessing.

For one thing, it is rarely pointed out that there are no specified developmental theories of ‘guessing’, or when and how guessing is to be expected as a response strategy. To be sure, there are discussions of various biases that might affect children’s performance on various tasks (such as a ‘yes’ bias; [Fritzley & Lee, 2003](#)). The advantage of theories that posit specific biases or response strategies is that they generate hypotheses that can be tested. Over the extent of the literature, they can be subjected to the same rigorous test as other theories of cognitive development (such as ‘conceptual change’, etc). Yet we know of no theoretical formulations that propose when and how children are disposed to generate random responding. Because the details of guessing theories are never articulated or specified, they become a convenient ‘catch-all’ theory—but a theory nonetheless that has no support from independent evidence. There does not seem any good reason to us to prefer such a theory to an articulated, testable one.

The idea that performance must be above chance to be meaningful is perhaps more usefully applied in cases where there is no background bias toward one or other response. It is notable in the false belief task that children rarely respond randomly in standard versions of the task—indeed the fuss in this domain has mostly been because children make such systematic errors. So if responding at a level that would not exceed chance expectations is rarely seen in such tasks, on what basis is it a benchmark for determining that a novel result is important? Part of the reason may reflect the tendency within studies of cognition in young children to frame successes and failures in terms of the child ‘having’ or ‘not having’ understanding of a concept. The idea that about half the children at a given age might get an answer right does not translate easily into this framework. One is tempted to worry whether they ‘get it’ or not.

These problems do not arise if one considers ‘the child’ and ‘understanding’ as divisible entities. A manipulation might have some systematic effect on the functioning of one of the processing elements that contributes to performance, leading to a systematic shift in the likelihood of a given response (see, e.g. [Scholl & Leslie, 2001](#)). Thus, a manipulation that moves children away from a specific error or pattern of performance by some non-trivial and statistically significant margin need not require that the child (as a whole) needs

now to be thought of as ‘understanding’ the concept, where previously she did not. Instead, the manipulation has simply had an effect on one aspect of the information-processing procedures that the cognitive system performs during the execution of the problem.

Turning to the current results, there are several reasons to suspect that the improvement in the youngest children in Experiments 1 and 3 was not induced by the manipulation creating false positives by inducing bias toward the correct location in response to the look first question.

First, all the children in Experiment 3 passed the true belief screening task prior to being admitted into the experiment. This gives us *prima facie* evidence to suppose that they were not susceptible to being biased by the manipulation, nor confused by it. The true belief tasks are a powerful control against the notion that manipulations induce false positives because of bias and or confusion. This is because the notion of a manipulation *biasing* children toward one response requires that the bias or confusion holds across different versions of that task. If not, then the bias account must somehow explain how the children know *not* to be biased on some of the occasions, but fail to know this on others. Demonstrations of ‘early’ false belief understanding via gaze direction or looking time measures (e.g. Clements & Perner, 1994; Onishi & Baillargeon, 2005) have made extensive use of true belief control conditions to show that low level response biases do not account for successes that are demonstrated in false belief detection tasks.

Secondly, the results fit within a broad and increasing literature of demonstrations of successful false belief reasoning in younger 3-year-old children, which we review briefly below. While isolated results demonstrating success on false belief tasks that do not demonstrate performance in excess of ‘chance’ might be conservatively attributed to guessing, as the number of demonstrations increases, it becomes increasingly plausible to suggest that such performance shifts reflect something real in terms of their effects on the child’s cognitive system. The notion that only manipulations that create ‘full understanding’ (assessed as better than chance responding) in ‘indivisible’ children, in our view, itself becomes the less conservative position.

### 5.2. Integrating ‘conversational’ and ‘executive inhibitory’ performance factors

It has been suggested that ‘conversational understanding’—thought of as a resource for computing answers on false belief tasks—becomes more efficient across development. This developmental process has been advanced as an explanation for 3-year-olds’ difficulty in false belief tasks—lack of success being attributed to resource limitations on children’s ability to acknowledge the experimenter’s purpose in asking the test question in an experimental setting (e.g. Siegal, 1997, 1999; Siegal & Varley, 2002). In this framework, the ‘look first’ question acts as a conversational aid, providing a performance resource that helps 3-year-olds to process the appropriate intent behind the experimenter’s question and calculate the correct response. This idea has great intuitive appeal, because the manipulation can be captured as one that ‘clarifies’ things for the younger child, whose conversational resources are less sophisticated than those of the older child. The older child, in possession of greater conversational resources, does not need the clarification in the standard version of the false belief task.

However, given a commitment to the computational theory of mind (see e.g. Fodor, 1975; Pinker, 1997), it is desirable to advance explanations of constructs such as ‘conversational resources’ and ‘clarification’ in terms of the framework theory of information processing. It is therefore instructive to make an attempt at capturing the effect of look first in terms of models of the processing of information during successful belief desire reasoning. Leslie and colleagues have advanced the only models of belief-desire reasoning in which the processes leading to successful performance have been specified (Friedman & Leslie, 2004a, 2004b, 2005; Leslie et al., 2004, 2005; Leslie & Pollizi, 1998).

The primary way in which success and failure on false belief tasks has been interpreted under these models is in terms of manipulations affecting the ability to control selection of responses via inhibition (see also German & Leslie, 2000, 2001; German & Nichols, 2003; Leslie, 1994a, 2000; Leslie & Thaiss, 1992; Roth & Leslie, 1998; Surian & Leslie, 1999).<sup>3</sup> Under this account, and following Surian and Leslie (1999), we can offer a mechanism through which look first might play its clarification role.

The idea is that the look first question draws attention to the first location (i.e. the target of the false belief content) and renders it more salient than it otherwise would be with standard questioning. The question format thereby tends to reduce the salience *differential* between true belief (default) and false belief contents. The reduced differential in turn requires less inhibition to reverse its direction, allowing children with lower inhibitory resources a greater chance of success (see Leslie et al., 2005). Note that the inhibition reduction account of look first enhances Siegal and Beattie’s (1991) account in providing a specific mechanism for how clarification might work (Surian & Leslie, 1999). Such a mechanism has greater explanatory scope because it not only explains how clarification can occur for the 3-year-old, but also why 4-year-olds do not need the same to help in order to pass standard false belief prediction.

Extending this account, we note that the same general explanation can be applied to explain why, even though 4-year-olds do not need help in the standard approach desire versions of false belief, they have difficulty in versions of the false belief problem with additional inhibitory selection demands such as where the action resulting from a false belief coupled with a desire to *avoid* the object must be predicted (Cassidy, 1998; Friedman & Leslie, 2004a, 2004b; Leslie et al., 2005; Leslie & Pollizi, 1998). In ‘avoid-desire’ versions of false belief action prediction, Sally’s desire is to avoid the target object (rather than approach it), and this small change actually results in problems among 4-year-old children who otherwise succeed at calculating false belief (they correctly solve a think question about Sally’s belief) and who can predict action if Sally wants to approach the object (Leslie & Pollizi, 1998). In fact, children do not reliably predict action from the combination of false belief and avoid desire until age 6 (Friedman & Leslie, 2004a).

Leslie and colleagues advance the idea that the avoid desire false belief task is additionally difficult for children because a second inhibitory process resulting from the

---

<sup>3</sup> We restrict this claim to models of tasks assessing belief-*desire* reasoning (e.g. action prediction false belief tasks such as the Sally-Ann task). Barreau and Morton (1999) offer a plausible memory-based model for performance in a task requiring belief attribution only (to the self).

avoid desire must be combined with the inhibitory process required to select the false belief content. These two processes must interact in such a way as to result in the prediction that search will take place at the location containing the object (which is the correct location, given that there are just two locations, Sally's false belief, and her desire to avoid the object).

Note that without the framework for conceptualizing performance factors in terms of information processing and inhibitory control, it is not obvious why 4-year-olds who have sufficient conversational resources to 'get the point' of the experimenter's question for tasks with approach desire should suddenly have problems for the very same question asked about a task with avoid desire. However, by conceptualizing 'conversational resources' in terms of effective control of inhibition we can make sense of this otherwise puzzling finding.

As noted earlier, this processing model also makes the prediction that look first might provide relief for the 4-year-old children who have difficulty with avoid desire, just as it provides relief for those 3-year-olds who otherwise would err in approach desire tasks. This is indeed the pattern of results that is seen (see Leslie et al., 2005), and moreover is a result that provides insight into determining which of the two different processing models of success proposed by Leslie and Pollizi (1998) are correct (see also Friedman & Leslie, 2004a, 2004b; Hehman & German, in press; Hehman et al., 2005).

The current paper is one of several recent attempts to integrate studies of the effects of various aspects of language development on the expression of mental state reasoning abilities. The interactions between language and 'theory of mind' comprise effects of syntactic development (Cheung et al., 2004; DeVilliers & DeVilliers, 2000; DeVilliers & Pyers, 2002), semantic development (Bloom, 2000; Happé & Loth, 2002;) and conversational pragmatic development (Woolfe, Want, & Siegal, 2002). While these approaches might be seen as importantly different from one another in their emphasis of different performance factors in explaining the pattern of mental state reasoning across the preschool period and beyond (see Leslie et al., 2005, for discussion), there is important common ground that can be explored in attempting to capture them in a common framework, and thereby bring them together also with work demonstrating relationships with domain general measures of executive function (e.g. Carlson et al., 1998, 2002), and studies of the relationship between other aspects of language use and theory of mind (Cheung et al., 2004; Hale & Tager-Flusberg, 2003; Lohmann and Tomasello, 2003; Siegal, Varley, & Want, 2001; Slade & Ruffman, 2005).

### 5.3. *Epilogue: the truth, the whole truth and nothing but the truth about false belief?*

Wellman et al.'s (2001) meta-analysis of false belief task performance was under the subtitle of 'The truth about false belief'. The meta-analysis is certainly informative, and is clearly a service to the field, but it is important not to over-extend the statistical conclusions that one is entitled to draw from such an enterprise. For example, the analysis shows that there is an evidence for improvement in the performance at false belief across the 3–5 year period, a statistical finding that is extended to a theoretical conclusion—that this change is best explained in terms of conceptual change.

As pointed out by commentators on the analysis (Moses, 2001; Scholl & Leslie, 2001), while the extensive evidence presented in the meta-analysis is *consistent* with conceptual change (e.g. there is variance in performance that is not explained by the known performance factors introduced into the ‘best’ model identified by Wellman et al., 2001), this does not actually demonstrate conceptual change. Moreover, this finding is not *inconsistent* with ‘early competence’ accounts, as look first exerts a significant effect even on the false belief performance of youngest 3-year-olds, and across different cultural settings.

The present series of studies has tested one major specific prediction of conceptual change accounts for one of the performance factors discussed in the Wellman et al. (2001) meta-analysis—the effect of ‘temporal marking’ (Freeman & Lacoheé, 1995; Lewis & Osborne, 1990; Siegal & Beattie, 1991), which included the look first task. This prediction, based on the output of the meta-analysis, was that only children toward the older age end of the 3–5 year range would benefit from temporal marking—children who might be thought of as in a transitional period between one theoretical understanding of mind and the next (Gopnik, 1996; Gopnik & Wellman, 1992, 1994). However, our results undermine even the statistical conclusion generated from the meta-analysis; there is in fact an advantage for performance in response to look first over standard questions even among children in the youngest part of the 3–5 year age range. In all meta-analyses, one gets out only some version of what is put in. If there are important datasets excluded, as will be inevitable with a cut-off date, the meta-analysis will not reflect these datasets. The ‘truth’ about false belief, even in terms of a statistical description of performance (and irrespective of theoretical considerations), is thus at best an approximation that falls short of the whole truth.

We would also urge caution, following Scholl and Leslie (2001), in the interpretation of unexplained variance as providing support for a particular mechanism of development (e.g. conceptual change). Statistically, unexplained variance is just that—*unexplained*; it cannot *a priori* be claimed as support for one position over any other. It is our view that performance factors should be treated not as ‘artifacts’ that should be ‘removed’ in order to understand development in the domain of theory of mind, but rather as major features of information processing models of successful belief-desire reasoning (see Leslie et al., 2004). These factors are worthy of study and explanation in their own right, and in doing so, scholars may advance toward a truth about false belief that is believable.

## Acknowledgements

Acknowledgments: This paper was prepared with the support of the Iranian Ministry of Science, Research and Technology and a grant from the Leche Trust (AAY), as well as support from the UCSB Academic Senate (TPG), and a grant from the University of Northumbria (MD). Special thanks are due to Netalie Levi whose undergraduate project assisted in developing the studies reported here, to Veronica Guzzardi for drawing the pictures for Experiment 3, to Louisa Chan Egan and Danielle Truxaw for pilot testing of the picture task, and to Rosemary Firth, Jeanet Ingwersen, and Richard Parker for

collecting the data reported in Experiment 3. Thanks also to two anonymous reviewers for comments on a previous version of this manuscript.

## References

- Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a 'theory of mind'? *Cognition*, *21*, 37–46.
- 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.
- Bernstein, D. M., Atance, C., Loftus, G. R., & Meltzoff, A. (2004). We saw it all along: Visual hindsight bias in children and adults. *Psychological Science*, *15*, 264–267.
- Birch, S. A. J., & Bloom, P. (2003). Children are cursed: An asymmetric bias in mental-state attribution. *Psychological Science*, *14*, 283–286.
- Birch, S. A. J., & Bloom, P. (2004). Understanding children's and adults' limitations in mental state reasoning. *Trends in Cognitive Sciences*, *8*, 255–260.
- Bloom, P. (2000). *How children learn the meanings of words*. Cambridge, MA: MIT Press.
- Bloom, P., & German, T. P. (2000). Two reasons to abandon the false belief task as a test of theory of mind. *Cognition*, *77*, B25–B31.
- Carlson, M. S., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development*, *72*, 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., & Hix, H. R. (1998). The role of inhibitory processes in young children's difficulties with deception and false belief. *Child Development*, *69*, 672–691.
- Cassidy, K. W. (1998). Three- and four-year-old children's ability to use desire-and belief- based reasoning. *Cognition*, *66*, B1–B11.
- Cheung, H., Chen, H.-C., Creed, N., Ng, L., Wang, S. P., & Mo, L. (2004). Relative roles of general and complementation language in theory-of-mind development: Evidence from Cantonese and English. *Child Development*, *75*, 1155–1170.
- Clements, W. A., & Perner, J. (1994). Implicit understanding of belief. *Cognitive Development*, *9*, 377–395.
- DeVilliers, J. G., & DeVilliers, P. A. (2000). Linguistic determinism and the understanding of false beliefs. In P. Mitchell, & K. J. Riggs (Eds.), *Children's reasoning and the mind* (pp. 191–228). Hove, UK: Psychology Press Ltd, 191–228.
- DeVilliers, J. G., & Pyers, J. E. (2002). Complements to cognition: A longitudinal study of the relationship between complex syntax and false belief understanding. *Cognitive Development*, *17*, 1037–1060.
- Fischhoff, B. (1975). Hindsight does not equal foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance*, *1*, 288–299.
- Fodor, J. A. (1975). *The language of thought*. New York: Crowell.
- Fodor, J.A. (1998). *Concepts*. Cambridge, MA: MIT Press
- Freeman, N., & Lacochee, H. (1995). Making explicit 3-year-olds' implicit competence with their own false beliefs. *Cognition*, *56*, 31–60.
- Friedman, O., & Leslie, A. M. (2004a). Mechanisms of belief-desire reasoning: Inhibition and bias. *Psychological Science*, *15*, 547–552.
- Friedman, O., & Leslie, A. M. (2004b). A developmental shift in processes underlying successful belief-desire reasoning. *Cognitive Science*, *28*, 963–977.
- Friedman, O., & Leslie, A.M. (2005). Processing demands in belief desire rereasoning: Inhibition or general difficulty? *Developmental Science*, *8*, 218–225.
- Fritzley, V. H., & Lee, K. (2003). Do young children always say yes to yes-no questions? A metacognitive study of the affirmation bias. *Child Development*, *74*, 1297–1313.



- German, T. P., & Leslie, A. M. (2000). Attending to and learning about mental states. In P. Mitchell, & K. J. Riggs (Eds.), *Reasoning and the mind* (pp. 229–252). Hove, East Sussex: Psychology Press, 229–252.
- German, T. P., & Leslie, A. M. (2001). Children's inferences from 'knowing' to 'pretending' and 'believing'. *British Journal of Developmental Psychology*, *19*, 59–83.
- German, T. P., & Leslie, A. M. (2004). No (social) construction without (meta) representation: Modular mechanisms as the basis for the acquisition of an understanding of mind. *Behavioral and Brain Sciences*, *27*, 106–107.
- German, T. P., & Nichols, S. (2003). Children's inferences about long and short causal chains. *Developmental Science*, *6*, 514–523.
- German, T.P. & Hehman, J.A. (in press). Representational and executive selection resource in 'theory of mind': evidence from compromised belief-desire reasoning in elderly adults. *Cognition*.
- Gopnik, A. (1993). How we know our minds: The illusion of first-person knowledge of intentionality. *Behavioral and Brain Sciences*, *16*, 1–14.
- Gopnik, A. (1996). The scientist as child. *Philosophy of Science*, *63*, 485–514.
- 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.
- Gopnik, A., & Wellman, H. (1992). Why the child's theory of mind really is a theory. *Mind and Language*, *7*, 145–171.
- Gopnik, A., & Wellman, H. (1994). The theory theory. In L. A. Hirschfeld, & S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 257–293). New York: Cambridge University Press, 257–293.
- Grice, H. P. (1975). Logic and conversation. In P. Cole, & J. L. Morgan, *Syntax and semantics* (Vol. 3) (pp. 41–58). New York: Academic Press, 41–58.
- Hale, C. M., & Tager-Flusberg, H. (2003). The influence of language on theory of mind" a training study. *Developmental Science*, *6*, 346–359.
- Happé, F., & Loth, E. (2002). 'Theory of mind' and tracking speakers' intentions. *Mind and Language*, *17*, 24–36.
- Hehman, J. A., Truxaw, D., Wertz, A. E., Niehaus, J. L., Roarty, M. P., & German, T. P. (2005, April). Biases in belief-desire reasoning across the lifespan. Poster presented at the biennial meeting of the *society for research in child development*, Atlanta, GA, April, 2005.
- Joseph, R. M. (1998). Intention and knowledge in preschoolers' conception of pretend. *Child Development*, *69*, 966–980.
- Leslie, A. M. (1987). Pretense and representation: The origins of 'theory of mind'. *Psychological Review*, *94*, 412–426.
- Leslie, A. M. (1994a). Pretending and believing: Issues in the theory of ToMM. *Cognition*, *50*, 211–238.
- Leslie, A. M. (1994b). ToMM, ToBy, and Agency: Core architecture and domain specificity. In L. Hirschfeld, & S. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 119–148). New York: Cambridge University Press, 119–148.
- Leslie, A. M. (2000). 'Theory of mind' as a mechanism of selective attention. In M. Gazzaniga (Ed.), *The new cognitive neurosciences* 2nd ed. (pp. 1235–1247). Cambridge, MA: MIT Press, 1235–1247.
- Leslie, A. M., Friedman, O., & German, T. P. (2004). Core mechanisms in 'theory of mind'. *Trends in Cognitive Sciences*, *8*, 528–533.
- Leslie, A. M., German, T. P., & Pollizi, P. (2005). Belief-desire reasoning as a process of selection. *Cognitive Psychology*, *50*, 45–85.
- Leslie, A. M., & Polizzi, P. (1998). Inhibitory processing in the false belief task: Two conjectures. *Developmental Science*, *1*, 247–254.
- Leslie, A. M., & Thaiss, L. (1992). Domain specificity in conceptual development: Neuropsychological evidence from autism. *Cognition*, *43*, 225–251.
- Lewis, C., & Osborne, A. (1990). Three-year-olds' problems with false belief: Conceptual deficit or linguistic artifact? *Child Development*, *61*, 1514–1519.
- Lohmann, H., & Tomasello, M. (2003). The role of language in the development of false belief understanding: A training study. *Child Development*, *74*, 1130–1144.
- Mitchell, P., & Lacoohé, H. (1991). Children's early understanding of false belief. *Cognition*, *39*, 107–127.
- Moses, L. J. (2001). Executive accounts of theory of mind development. *Child Development*, *72*, 688–690.

- Nelson, K., Skwerer, D. S., Goldman, S., Heneseler, S., Presler, N., & Walkenfeld, F. F. (2003). Entering a community of minds: An experiential approach to 'theory of mind'. *Human Development, 46*, 24–46.
- Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science, 308*, 255–258.
- Perner, J. (1991). *Understanding the representational mind*. Cambridge, MA: MIT Press.
- Perner, J. (1995). The many faces of belief: Reflections on Fodor's and the child's theory of mind. *Cognition, 57*, 241–269.
- Perner, J., & Lang, B. (2000). Theory of mind and executive function: Is there a developmental relationship? In S. Baron-Cohen, H. Tager-Flusberg, & D. Cohen (Eds.), *Understanding other minds: Perspectives from autism* (pp. 155–186). Oxford: Oxford University Press, 155–186.
- Pinker, S. (1997). *How the mind works*. New York: Norton.
- Roth, D., & Leslie, A. M. (1991). The recognition of attitude conveyed by utterance: A study of preschool and autistic children. *British Journal of Developmental Psychology, 9*, 315–330.
- Roth, D., & Leslie, A. M. (1998). Solving belief problems: Toward a task analysis. *Cognition, 66*, 1–31.
- Scholl, B. J., & Leslie, A. M. (1999). Modularity, development and 'theory of mind'. *Mind and Language, 14*, 131–153.
- Scholl, B. J., & Leslie, A. M. (2001). Minds, modules, and meta-analysis. *Child Development, 72*, 696–701.
- Siegal, M. (1997). *Knowing children: Experiments in conversation and cognition* (2nd ed.). Hove, UK: Psychology Press.
- Siegal, M. (1999). Language and thought: The fundamental significance of conversational awareness for cognitive development. *Developmental Science, 2*, 1–14.
- Siegal, M., & Beattie, K. (1991). Where to look first for children's knowledge of false beliefs. *Cognition, 38*, 1–12.
- Siegal, M., & Varley, R. (2002). Neural systems involved in 'theory of mind'. *Nature Reviews Neuroscience, 3*, 463–471.
- Siegal, M., Varley, R., & Want, S. C. (2001). Mind over grammar: Reasoning in aphasia and development. *Trends in Cognitive Sciences, 5*, 296–301.
- Slade, L., & Ruffman, T. (2005). How language does (and does not) relate to theory of mind: A longitudinal study of syntax, semantics, working memory and false belief. *British Journal of Developmental Psychology, 23*, 117–141.
- Surian, L., & Leslie, A. M. (1999). Competence and performance in false belief understanding: A comparison of autistic and three-year-old children. *British Journal of Developmental Psychology, 17*, 141–155.
- Wellman, H. M. (1990). *The child's theory of mind*. Cambridge, MA: MIT Press.
- Wellman, H. M. (2002). Understanding the psychological world: Developing a theory of mind. In U. Goswami (Ed.), *Blackwell handbook of childhood cognitive development* (pp. 167–187). Oxford, UK: Blackwell, 167–187.
- Wellman, H. M., & Bartsch, K. (1988). Young children's reasoning about beliefs. *Cognition, 30*, 239–277.
- Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory mind development: The truth about false-belief. *Child Development, 72*, 655–684.
- Woolfe, T., Want, S. C., & Siegal, M. (2002). Signposts to development: Theory of mind in deaf children. *Child Development, 73*, 768–778.
- Zaitchik, D. (1991). Is only seeing really believing? Sources of the true belief in the false belief task. *Cognitive Development, 6*, 91–103.