Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Cognitive Psychology 50 (2005) 45–85
Cognitive Psychology
www.elsevier.com/locate/cogpsych

# Belief-desire reasoning as a process of selection

Alan M. Leslie[a,*], Tim P. German[b], Pamela Polizzi[a]

[a] *Department of Psychology and Center for Cognitive Science, Rutgers University, 152 Frelinghuysen Road, Piscataway, NJ 08854, USA*
[b] *Department of Psychology, University of California, Santa Barbara, USA*

Accepted 10 June 2004
Available online 22 September 2004

**Abstract**

Human learning may depend upon domain specialized mechanisms. A plausible example is rapid, early learning about the thoughts and feelings of other people. A major achievement in this domain, at about age four in the typically developing child, is the ability to solve problems in which the child attributes false beliefs to other people and predicts their actions. The main focus of theorizing has been why 3-year-olds fail, and only recently have there been any models of how success is achieved in false-belief tasks. Leslie and Polizzi (Inhibitory processing in the false-belief task: Two conjectures. Developmental Science, 1, 247–254, 1998) proposed two competing models of success, which are the focus of the current paper. The models assume that belief-desire reasoning is a process which selects a content for an agent's belief and an action for the agent's desire. In false belief tasks, the theory of mind mechanism (ToMM) provides plausible candidate belief contents, among which will be a 'true-belief.' A second process reviews these candidates and by default will select the true-belief content for attribution. To succeed in a false-belief task, the default content must be inhibited so that attention shifts to another candidate belief. In traditional false-belief tasks, the protagonist's desire is to approach an object. Here we make use of tasks in which the protagonist has a desire to avoid an object, about which she has a false-belief. Children find such tasks much more difficult than traditional tasks. Our models explain the additional difficulty by assuming that predicting action from an avoidance desire also requires an inhibition. The two processing models differ in the way that belief and desire inhibitory processes combine to achieve successful action prediction. In six experiments we obtain evidence favoring one model, in which parallel inhibitory

* Corresponding author.
*E-mail address:* aleslie@ruccs.rutgers.edu (A.M. Leslie).

0010-0285/$ - see front matter 
doi:10.1016/j.cogpsych.2004.06.002

processes cancel out, over the other model, in which serial inhibitions force attention to a previously inhibited location. These results are discussed in terms of a set of simple proposals for the modus operandi of a domain specific learning mechanism. The learning mechanism is in part modular—the ToMM—and in part penetrable—the Selection Processor (SP). We show how ToMM–SP can account both for competence and for successful and unsuccessful performance on a wide range of belief-desire tasks across the preschool period. Together, ToMM and SP attend to and learn about mental states.


*Keywords:* Belief desire reasoning theory of mind development inhibition domain-specificity

## 1. General introduction

Following Chomsky (1957, 1975), there has been growing interest in domain specialized learning mechanisms. Over the last 20 years it has become apparent that learning about the mental states of other people begins early, in the preschool years, and occurs even in those suffering moderate intellectual retardation (Leslie, 2000b). Given the highly abstract nature of mental states, such facts have suggested to many researchers that there is a specific innate basis to acquiring a 'theory of mind'—that is, a domain specialized learning mechanism.

One of the main proponents of domain specialized learning for 'theory of mind' has been Leslie and colleagues by way of the theory of the 'theory of mind' mechanism (ToMM) (e.g., Baron-Cohen, 1995; Frith & Frith, 1999; Gallagher & Frith, 2003; German & Leslie, 2001; Leslie, 1987, 1994a, 2000b; Leslie & Thaiss, 1992). According to this view, the representational system underlying 'theory of mind' comes on-line during the second year of life. Although 'theory of mind' is a much wider domain, a central component is the concept BELIEF which together with the concept DESIRE plays a key role in interpreting and predicting behavior. The principle developmental function of ToMM is to permit the child to attend to mental states that, unlike behavior, are otherwise invisible and undetectable. Once the child can attend to these states, and only then, she can begin to learn about them. Despite being able to detect mental states, the child still has everything to learn, because in this view these concepts are simply representations and do not depend upon having a theory about what mental states really are (Leslie, 2000a).

The ToMM theory of innate representations for mental states solves a number of difficult problems, such as learnability (Leslie, 2000b), provides an account of key aspects of the developmental disorder of autism (Frith, Morton, & Leslie, 1991), and is supported by findings from brain imaging studies (Frith & Frith, 1999). However, a major obstacle to the wider acceptance of this theory has been the findings from studies of the 'false-belief task,' in particular, that children typically fail until they are 4 years of age. Although the false belief task comes in several forms, dozens of studies have shown that performance on all of them is closely related (Wellman, Cross, & Watson, 2001). For reasons we will come to, we focus here on the 'most' standard of these tasks, namely, tasks derived from Wimmer and Perner's (1983)

Maxi task, such as the Sally and Ann task (Baron-Cohen, Leslie, & Frith, 1985), which the majority of 4-year-olds pass and the majority of 3-year-olds fail.

The main thrust of this paper is that solving a false-belief problem requires overcoming the default attribution of a belief that is true.[1] Overcoming the default true-belief attribution is necessary for success on false-belief problems. However, it requires an 'executive' selection process to *inhibit* the default attribution, and this selection process is relatively slow to develop. We will describe and test two specific models of this process. Both models incorporate the principles of (a) automatic attribution of true-belief and (b) selection of the false-belief by inhibition.

According to the above view, the failure of children on false-belief tasks is the result of a failure in the selection process to effectively inhibit the true-belief attribution. To the extent that this is true, there are a number of important theoretical consequences. First, what has been the single most important objection to the ToMM account is removed. Second, we will have advanced our understanding of a domain-specialized learning process. Attributing beliefs, with false *or* with true contents, is unique to the 'theory of mind' domain, yet children may come to learn about false beliefs by way of increased powers of selection by inhibition. Third, we will extend our understanding of the role of inhibition in reasoning and development. Thus, although belief attribution is domain specific, selection by inhibition may operate in a variety of domains. In this case, belief-desire reasoning and its development provide an avenue for studying inter-relations between domain specific and domain general processes of learning and reasoning.

### *1.1. The importance of success for modeling*

Despite intense focus on the question of when children first succeed in reasoning about false beliefs and why very young children fail (for a recent exchange, see Wellman et al., 2001, and Scholl & Leslie, 2001; see also Bloom & German, 2000), little attention has been paid to the question of *how* children *succeed* on false-belief tasks. Models of how tasks are successfully processed are ubiquitous in cognitive science but relatively rare in developmental research. Although the question of *when* the child first succeeds is interesting, more important is the question of *how* success is achieved, whenever that may be. There are several reasons for this. A model of success is a theory of *what* develops, *what* is learned. It is hard to see how development of $x$ could ever be understood without understanding what $x$ is. Furthermore, success is more constrained than failure. To succeed in building a house means meeting a set of demanding constraints (walls must bear weight of roof, etc); but one can *fail* to build a house in any number of ways. Likewise cognitive processing that regularly solves a task must meet demanding constraints. Constraints from success guide construction of theories of successful performance. Finally, a focus on failure can leave success unstudied: even if 3-year-old failure on false-belief is due to lack of an appropriate concept, it remains to be explained how that new-gained concept is

[1] Because we are discussing belief attribution, the truth or falseness of a belief always means 'from the point of view of the attributer.'

employed to produce success. Passing the false-belief task shows that the child has the concept BELIEF but it does not tell us how the task is passed. Likewise, successful recall tells us that someone remembered something, but it does not tell us *how* memory *works*.

Our aim is to model how information in false-belief tasks is processed for success. The typical 4-year-old can correctly predict the behavior of another person when that person has a false-belief. However, this is only true when the other person's desire is to *approach* the target about which they have the false-belief. Surprisingly, if the protagonist desires to *avoid* the target, then 4-year-olds perform just as poorly as 3-year-olds in predicting behavior (Cassidy, 1998; Leslie & Polizzi, 1998).

For Leslie and Polizzi (1998), the above finding provides support for the 'selection processing' (SP) model, first put forward in Leslie and Thaiss (1992; see also German & Leslie, 2000; Leslie & Roth, 1993; Roth & Leslie, 1998). The basic idea of selection processing is to select the most plausible belief content from among a small set of plausible candidates. The candidates are provided by an automatic, and possibly modular, process that runs when we attend to the behavior of an agent. This automatic process—associated with the theory of mind mechanism (ToMM)—attributes to the agent relevant beliefs and desires (Leslie, 1987, 1991, 1994a; for a recent review, see Leslie, 2000b).

Below we examine in more detail the general idea of selection processing. This leads to a discussion of each of Leslie and Polizzi's two models for how selection processing occurs in simple belief-desire reasoning. We then present a series of experiments that test key assumptions of both models. We confirm the basic avoidance-desire effect on false-belief tasks and clarify the basis of the effect, showing that a desire that is specified negatively is neither necessary nor sufficient for producing the effect. Instead, we argue that the critical factor is the pattern of inhibitory processing in a belief-desire task that requires shifts in the target of attention. We then investigate whether the 'look first' manipulation that helps 3-year-old failers to pass standard false belief (Siegal & Beattie, 1991; Surian & Leslie, 1999) can help 4-year-olds to pass false-belief tasks that require complex inhibition. We find that children that reliably pass single inhibition false-belief tasks are helped by a 'look first' question to pass double inhibition tasks. Finally, we consider grounds for preferring one of the two models over the other.

### *1.2. Selection processing in the false-belief task*

Selection processing (SP) has been conceptualized as an executive or control process necessary to select among competing alternative representations. Usually, selection does not pose much of a problem because ToMM always offers a 'true-belief,' that is, a belief with a content that is true (in the eye of the attributer). A true-belief is always more highly valued by SP and is selected by default. A true-belief default is ecologically valid because, at least about mundane matters, people's beliefs usually *are* true. We can go a little further than this. For a basic belief-attributing system—one whose business concerns simple everyday beliefs—the true-belief attribution *ought* to be the default. This is because, in the absence of specific information, the only general constraint on belief attribution is provided by the state of the world (as it appears to the attributer). For a given marble on a given occasion, a true-belief must refer to its real location, but indefinitely many locations may be mentioned by false beliefs. Given that fact, in the absence of specific information to the contrary, a belief attributer's best guess will always be the true-belief.

Although, on average, a default selection of the true-belief will be correct more often than incorrect, in false-belief tasks the default content is guaranteed to be wrong and a more complex selection process must be employed. In a standard false belief task, a protagonist, Sally, hides a marble in a basket which unbeknownst to her is then moved to a box. According to the ToMM–SP model, ToMM spontaneously identifies (at least) two possible contents for Sally's belief, namely, that *the marble is in the basket* (false-belief content) and that *the marble is in the box* (true-belief content). These candidate contents are then reviewed by the selection processor. SP is designed such that the true-belief content is 'prepotent' or 'salient' relative to the false-belief content. Consequently, in a false-belief task, selection of the correct false-content requires that the (incorrect) true-content must first be rejected. Rejection of the default may be achieved by inhibition.

The rejection-by-inhibition hypothesis was put forward in part because it is consistent with a larger executive function framework. However, selection processing can be formulated in a number of different ways than that adopted by Leslie and Polizzi (1998). For example, we could say that the true-content is more highly 'activated' than the false-content. Or we could say that the true-content has a higher 'subjective probability' (of being correct) than the false-content. Then instead of inhibition we could speak of 'lowering the activation level' or 'decreasing subjective probability.' Or we could say that attention has positive and negative polarities. At this point, as far as we know, these are all equally valid. We are therefore in principle neutral at this point between these different formulations. However, we will adopt the terminology of Leslie and Polizzi (1998) and speak throughout of relative 'salience' and 'inhibition.'

The common thread that runs through all the formulations is as follows. The typical mode of operation (MO) for ToMM is to offer more than one candidate content for a mental state attribution. There is a default *preference differential* between true-belief and false-belief contents, favoring the true-content. Selection processing reviews the candidates on offer and may modify their initial preference levels in light of specific circumstances. The most highly preferred candidate at the end of this process becomes the belief that ToMM–SP finally attributes to the protagonist.

### 1.3. Two inhibition models of selection processing

Two main findings form the basis for Leslie and Polizzi's development of the ToMM–SP model. First, the apparently minor change of giving the protagonist a desire to *avoid* rather than approach a target, makes 4-year-old subjects, who pass a standard false belief task, perform at a thoroughly 3-year-old level. Leslie and Polizzi (1998) and Cassidy (1998) found that of these children only 37.5 and 38%, respectively, were able to pass. Since these children comprehend both the false-belief and the avoidance desire, difficulties of a conceptual nature cannot be the issue. A performance account is needed.

The second finding was a large divergence in performance between two questions routinely used to assess false-belief understanding. The 'Think' question asks where Sally *thinks* the marble is, whereas the 'Prediction' question asks where Sally will *look for* the marble. 'Think' questions require only an ascription of belief, while a prediction of behavior question requires an additional ascription of desire, the integration of belief with desire, and the inferring of a resulting action. It is striking, then, that in standard tasks—approach-desire + false-belief—there is no measurable effect of the additional processing for Prediction over Think. Extensive data gathered over the last 15 years shows an almost perfect correlation between children's performance on the two questions in standard false belief tasks, a finding confirmed by Wellman et al.'s (2001) recent meta-analysis. By contrast, in an avoidance-desire + false-belief task, Leslie and Polizzi (1998) found a wide divergence between performance on Think and Prediction questions (100% versus 38% passing).

To account for the dramatic effect of an avoidance desire on 4-year-old performance, Leslie and Polizzi introduced the notion of a *double inhibition*, and described two different ways in which double inhibition could produce the observed effects. Both models extend and develop the idea of selection processing; and both are models of successful performance. Both characterize a review and selection process that adjusts initial salience levels; and both account for the same range of belief-desire reasoning tasks, namely, the (easy) approach-desire + true-belief task, the (standard) approach-desire + false-belief task, the (easy) avoidance-desire + true-belief task, and the (difficult) avoidance-desire + false-belief task.

The models can be described by imagining that the belief metarepresentation computed in the Sally–Ann task contains a variable *P* in the decoupled content slot, **Sally believes (of) the marble (that) *"P"***. Candidate values of this variable, **"it is in the box," "it is in the basket,"** are 'displayed' simultaneously. Selection is determined by a (mental) pointer that is attracted to the more salient of the candidate contents. To achieve success in some tasks, the initial salience level of a candidate needs to be adjusted. This is done by applying inhibition to that candidate, lowering its salience level. If, as a result of inhibition, a candidate drops in salience below that of the alternative, then the mental index will swing across to the now higher valued candidate. If no further adjustments are made, then the currently indexed candidate will be bound to *P* and selected as best guess regarding Sally's belief. When, as a result of applying inhibition, the index swings from one candidate to another, we will say that a *target shift* has occurred.

The notion of a mental index is useful because it helps us visualize the child's task as determining which of two locations is referenced by Sally's belief, desire, or predicted action. We can also think of the index as standing for the child's attention. Shifting attention from one target to another is a ubiquitous psychological process and has been intensively studied in the case of vision. The shifting of visual attention is widely believed to require an inhibitory process that disengages attention from its current target before it can be moved to another (e.g., Posner & Cohen, 1984; Rafal & Henik, 1994). If either of our models is correct, then shifting visual attention and shifting attention in reasoning may have underlying similarities.

We assume that ToMM offers only *plausible* contents to SP. Therefore, there will be only two candidates on offer in a standard false-belief task, namely, the last location of the bait to which the believer had access and the current location of the bait.

Where Leslie and Polizzi's models differ from one another is in depicting how belief and desire selections are made—either in parallel or serially. In Model 1, the *inhibition of inhibition* model, belief and desire are identified in parallel, and inhibitions, where appropriate, are applied in parallel. Because these operations are performed in parallel, only a single index is required. In a task in which belief alone needs to be identified (e.g., for a Think question) or where desire alone needs to be identified (in a desire task), a single index is obviously sufficient. But where belief and desire need to be considered together, as in a Prediction task, then the single index must do double service. A single index may minimize processing demands, but it means that the target of the protagonist's desire must be identified directly in relation to the world (and not in relation to belief).

Model 2, the *inhibition of return* model, takes the more sophisticated approach of identifying targets serially, with belief targets identified first and desire targets second. This requires two indexes to be employed, one for belief and one for desire. Perhaps this is more demanding computationally. It does, however, allow the target of desire to be identified in relation to the protagonist's belief (about the world), which seems a more 'adult' mode of operation.

These differences in how belief and desire targets are identified—in parallel or serially, using combined or separate indexes, directly or indirectly—have a number of consequences for the successful processing of false-belief tasks. One advantage of specifying these models, even informally, is that we can more easily draw out empirical consequences and theoretical properties that are not at first obvious. We now discuss each of the models in turn.

### 1.3.1. Model 1: Inhibition of inhibition

Fig. 1 graphically illustrates how this model works for four different tasks. The index is shown as a pointing hand, and the alternative belief contents/targets as boxes. Inhibition is shown as a red arm that can grasp a content/target with the effect of inhibiting it, reducing its salience. A matrix shows the MO of SP across a number of tasks. Since beliefs can be true or false and desires for approach or avoidance, four possible tasks are shown with beliefs (true/false) in the columns and desires (approach/avoidance) in the rows.

The first cell shows the simple true-belief + approach-desire task (e.g., Sally wants the marble and knows where it is). In Model 1, because only one index is used, the target of belief and of desire are necessarily identified in parallel. The true-belief target is indexed initially, and because the target of desire is identified at the same time, it too initially points to the target where the bait actually is. This task is very simple because neither the initial identification of belief nor desire targets needs subsequently to be adjusted.

The next cell shows the (standard) false-belief + approach-desire task (Sally wants the marble but has a false-belief about its location). Again the target of true-belief is initially indexed but this time it is subsequently inhibited because the belief is false.

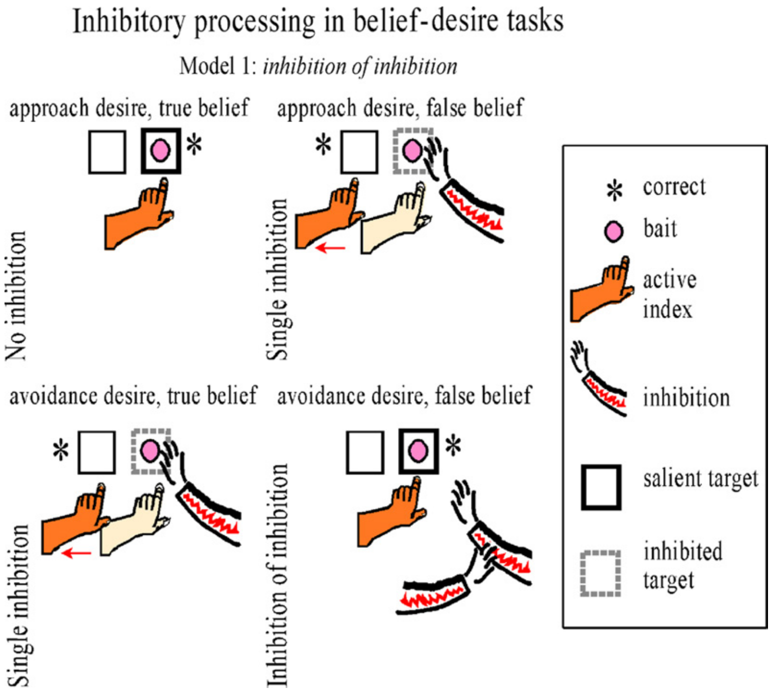


Fig. 1. A model of selection processing in belief-desire reasoning. The panels are arranged to illustrate a $2 \times 2$ factorial design with approach/avoidance desire in the rows and true/false-belief in the columns. The pointing hand represents a mental index indicating the target of belief or of belief-desire and thus the answer to a 'think' or a 'prediction of behavior' question, respectively. The grabbing arm represents an inhibitory process that reduces the salience of the target to which it is applied. The target of true-belief is initially more salient but can be subsequently inhibited if the belief is false (second panel) or the desire is to avoid (third panel). Weakening of a target causes the index to move to the alternate now more salient target. The final panel shows 'double inhibitions' canceling out, rather than summing, to give the correct answer to false-belief with avoidance-desire problems. (After Leslie & Polizzi, 1998.)

Since the desire is for approach, no desire inhibition is generated. The belief inhibition causes the index to swing across to the false-belief target; consequently it is selected. For this task, Model 1 operates in exactly the same way for both a Think question and a Prediction question.

The next cell of Fig. 1 shows a true-belief + avoidance-desire task (Sally knows where the marble is but wants to avoid it). Again the target of true-belief is initially indexed but this time it is inhibited not by a belief inhibition but by a desire inhibition (the target initially identified is exactly what Sally does *not* want). Again the index swings across, this time selecting the correct target of desire.

The final cell shows a task in which an avoidance-desire is coupled with a false-belief (Sally wants to avoid the marble but has a false-belief about its location). As always, the target of true-belief is initially identified. In this task, however, two inhibitions are generated, one because the belief is false and the other because the desire is to avoid. Notice, however, that the two inhibitions cannot simply be applied in combination. If they were, they would force the index to swing across to the alternate target. But this is the wrong answer for this task and we want to model success, not failure. Instead of summing the two inhibitions, they must cancel out, that is, inhibit *each other*. Because no inhibition reaches the target, there is no shift: the index stays put.

It is reasonable to assume that marshaling an inhibition of inhibition is considerably more difficult than applying one or more single inhibitions to a target. In avoidance false-belief tasks, failure to control inhibition of inhibition while successfully marshaling single inhibition will produce correct answers to the Think question along with incorrect answers to the Prediction question. This response pattern is typical of 4-year-olds (Leslie & Polizzi, 1998).

#### *1.3.2. Model 2: Return to inhibition*

Fig. 2 graphically illustrates how this model works for the same four tasks. In this model, separate indexes are used for belief and desire, and are processed serially (belief first). Once again, the target of true-belief is identified first. Any inhibition necessary to produce a target shift (if the belief is false) is applied to the belief index (second and fourth cells). The target of desire is then subsequently identified relative to the final target of belief. For example, in a standard task (second cell), the target of belief is finally identified as the empty location; the target of Sally's desire is identified relative to *that*—the belief that the marble is in the basket. Notice that in this model, the addition of the desire index is required only by the Prediction question. Unlike the first model, Model 2 works slightly differently for the Think and Prediction questions. Having initially identified desire in relation to belief, any inhibition necessary to produce a desire target shift is then applied (if the desire is to avoid—third and fourth cells). The final placement of the desire index determines the target of action (Prediction question). The difficulty of the avoidance-desire + false-belief task (fourth cell) is accounted for in terms of a previously inhibited target resisting the return of an index (attention). This is reminiscent of an effect in the visual attention literature, known as "inhibition of return," in which attention resists return to a previously inhibited target (Posner & Cohen, 1984).

The second model takes a more sophisticated approach than the first model by identifying desires in the light of beliefs. However, in implementing this more advanced approach, the second model must employ two indexes and process them serially. The first model makes do with a single index but then must identify belief and desire in parallel.

What *triggers* the inhibition is an interesting question but not one we will address here. Presumably, the recognition that Sally does not see/know that the bait is in the new location plays a role. Because the models capture successful performance, inhibitions, howsoever triggered, must be strong enough to produce a target shift.

### *1.4. Divergence between belief attribution and behavior prediction*

Although the models under discussion differ only subtly in their details, it should be possible to distinguish them empirically. The extra step of adding the desire index

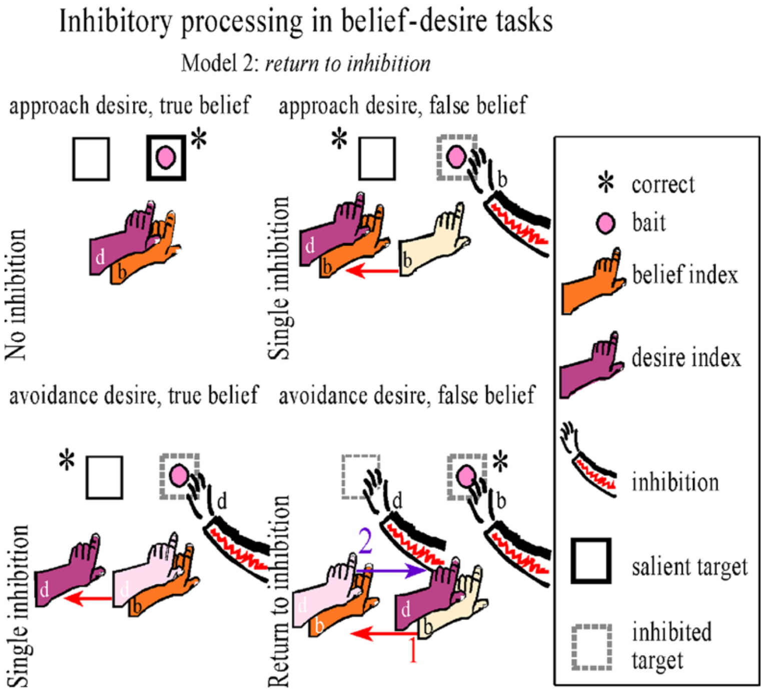


Fig. 2. An alternative model of selection processing in belief-desire reasoning. In this model belief-targets are identified first and desire-targets second and in relation to the identified belief-target. Again, the true-belief target is initially more salient and thus indexed but, subsequently, the belief-target is inhibited if the belief is false (second panel) and the desire-target inhibited if the desire is to avoid the target (third panel). The final panel shows the resulting sequence in the 'double inhibition' task. First, the target of true-belief is identified and inhibited, causing the belief index to move to the alternative target. The target of desire is then initially identified in relation to the (false) belief target but then inhibited (for avoidance). The desire index is then forced to return to the true-belief target, despite the residual inhibition there. (After Leslie & Polizzi, 1998.)

to answer the standard Prediction question (empirically without cost) does not compellingly disfavor Model 2, because a Model 2 theorist can assume that adding that index has a processing cost too small to measure.

However, an adequate model must also account for the fact that, in the 'double inhibition' task, 4-year-olds pass the Think question and then immediately fail Prediction. Apparently, children cannot simply remember the attributed false-belief for a moment while they add in the desire. We know that adding an avoidance-desire to a *true*-belief poses negligible difficulty for these children—almost all pass (Leslie & Polizzi, 1998). So, with false-belief already calculated, with the price of figuring out that answer *already paid*, why should it be dramatically harder to add an avoidance-desire?

We saw earlier that the models give different accounts of the interaction between false-belief and avoidance-desire. However, now we are asking: Given that false-belief has *already* been calculated in response to the Think question, why, for Prediction, should false-belief processing *still* interact with that of avoidance-desire? Again, the models give different answers.

### 1.4.1. Inhibition of inhibition

With this parallel model, both belief and desire targets must be processed simultaneously because there is only one index. This means that to answer the Prediction question, there is no choice but to calculate belief as well. Even if belief has already been correctly identified to answer the Think question, it must be calculated all over again to correctly answer the Prediction question. We shall refer to this property of Model 1 as *mandatory recalculation*.

Couldn't false-belief be identified first, in response to the Think question, and then, for a closely following Prediction question, couldn't desire-processing start with the index already at the false-belief target? Couldn't desire-inhibition be applied there, shifting the index back to the full location and yielding the correct answer? Yes, it could; this even seems more 'sensible.' However, this process describes not Model 1, but Model 2.

### 1.4.2. Inhibition of return

This model can account for the lack of savings from answering Think before Prediction without any recalculation. If false-belief has already been calculated (for the Think question), the target of true-belief will already be inhibited and the target of false-belief (empty location) already indexed. Prediction will then proceed with identification of approach-desire relative to (false) belief, followed by (avoidance) inhibition. However, the inhibition (from answering the Think question) lingers on the full location, making return there difficult. (Note that lingering inhibition is critical to the visual attention version of inhibition of return.) Model 2 thus accounts for lack of savings from Think to Prediction without requiring recalculation.

## 1.5. Structure of the experiments

Our main aim is to explore and test the above models, both the general principles that unite them and some of the points that distinguish them. We report a series of experiments that test each of the cases represented by the different cells in Figs. 1 and 2. The only exception to this is the first cell which represents a task in which Sally knows where the marble is and wants it. We felt this task was so simple that we would risk insulting our subjects. The other task cells are a standard false belief task, with which we routinely screen subjects, avoidance true belief and avoidance false-belief. The latter two tasks form the focus of our investigations. Although we use unexpected location tasks throughout, we believe that our models apply generally to false-belief tasks. To probe the nature of belief processing, we exploit desire and action prediction. Our reason for using the unexpected location task is that it incorporates desire and action prediction in a more straightforward way than, for example, deceptive appearance tasks.

The first experiment checks the reliability of the basic experimental findings on avoidance desire by attempting to replicate Leslie and Polizzi (1998) in a different laboratory. The second experiment asks whether the double inhibition effect can also be produced by opposite pretending. The third experiment shows that 'target-shifting,' rather than complexity, is critical to recruiting inhibition for avoidance desire. The fourth experiment shows that for very young children a desire involving a 'target-shift' is harder than a desire that does not. The final two experiments probe whether a task manipulation known to make the standard false belief task easier for 3-year-olds also makes the double inhibition task easier for 4-year-olds. The results favor one of our models and suggest that selection by inhibition operates in the belief processing of both 3- and 4-year-olds.

## 2. Experiment 1

We attempted to replicate the findings of Leslie and Polizzi (1998).

### 2.1. Methods

#### 2.1.1. Subjects

Thirty-six children were seen. Subjects were required to pass a standard false-belief test based on the 'Sally and Ann' task of Baron-Cohen et al. (1985). Two children were excluded from further testing for failing this screening task. The remaining 34 subjects (17 boys) were aged between 4 years 0 months and 5 years 7 months (mean age = 4 years 8 months, $SD = 5$ months), with 17 subjects randomly assigned to each of two groups. Children were recruited from and tested in quiet areas in schools in Essex, England, and were drawn from diverse SES and were predominantly Caucasian.

#### 2.1.2. Materials

Materials included three toy rooms constructed from cardboard, one for each of the tasks (including screening task), distinctly colored boxes, and small dolls and props used to enact scenarios.

#### 2.1.3. Design and procedure

We presented two tasks in story form, Avoidance Desire and Opposite Behavior. Each task was presented in both true and false-belief versions, yielding four conditions. Group 1 received the Avoidance Desire True Belief story and Opposite Behavior false-belief story. Group 2 received the Avoidance Desire false-belief story and Opposite Behavior True Belief story. Thus, each subject was randomly assigned to two of the four conditions with the constraint that no child received both true and false-belief versions of the same story. The story protocols used were identical to Leslie and Polizzi (1998, See Appendix).

*2.1.3.1. Avoidance desire task.* A girl was described as not wanting to put food in a box containing a sick kitten, otherwise the kitten would eat the food and become worse. In the true belief condition, the girl watched the kitten move from box A to box B. In the false-belief condition, she observed the kitten in box A but was absent when it moved to box B.

*2.1.3.2. Opposite behavior task.* A "mixed-up man" was described as always doing the opposite of what he desires. If an object is in box A, he would look for it in box B. In the true belief condition, he watched as his Mexican jumping bean jumps from box A to box B, while in the false belief condition, he was absent as it moved.

Subjects who failed Memory or Reality (control) questions were corrected the first time, the story was retold up to that point and the control question asked again. A second failure would have meant rejection but in fact no child failed a control a second time. To maintain pragmatic naturalness, subjects were asked a Know question in true belief conditions and a Think question in false-belief conditions. Departing from Leslie and Polizzi (1998), we treated the Know and Think questions as control questions and adopted the same procedure as for the Memory and Reality questions. Subjects who failed were corrected, recycled through the story, and asked the Know or Think question again. No subject failed these questions a second time. All subjects were asked a Prediction question.

In true belief conditions, passing requires indicating the location opposite to where the object is in reality. In the false-belief conditions, passing requires indicating the box where the object actually is. Better performance in true belief than in false-belief conditions was predicted.

### 2.2. Results

Fig. 3 shows that 94% of subjects passed Prediction in the true-belief version of the Avoidance Desire story while only 12% passed the false-belief version (Upton's $\chi^2 = 22.5$, $p < .001$, one-tailed).[2] A similar pattern was observed on the Opposite Behavior story: 82.4% of subjects passed Prediction in the true-belief version while only 12% passed the false-belief version (Upton's $\chi^2 = 18.3$, $p < .001$, one-tailed). In the Avoidance Desire groups, one subject failed the Know question first time round, and two subjects failed the Think question first time round. We re-analyzed the results excluding these children. The same pattern was found: 94% passed true-belief and 15% passed false-belief (Upton's $\chi^2 = 19.6$, $p < .001$, one-tailed). In the Opposite Behavior groups, no child failed the Know question and three children failed the Think question. Re-analyzing the data with these children excluded showed the same pattern with 82.4% passing true-belief and 8% passing false-belief (Upton's $\chi^2 = 16.8$, $p < .001$).

#### 2.2.1. Discussion

These results are in close agreement with Leslie and Polizzi (1998). Specifically, we confirmed that, despite correctly attributing a false belief to the protag-

[2] For Upton's $\chi^2$ see Richardson (1990).

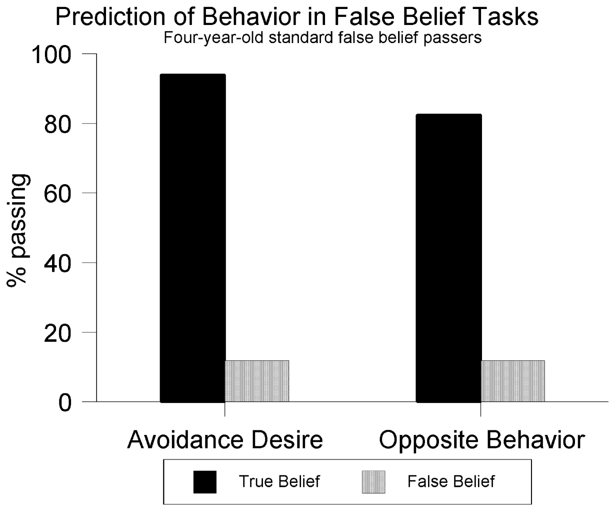


Fig. 3. Experiment 1: Children who pass standard false belief scenarios (with an approach desire) perform poorly when the desire is to avoid or if the scenario character is disposed to 'opposite' behavior.

onist in answering the Think question, 4-year-olds cannot then correctly predict the protagonist's behavior. Furthermore, they fail at this both when avoidance desire and when opposite behavior must be combined with the false-belief attribution.

There appear to be a number of ways to produce what we call 'target-shifting.' So far, these are attributing a false-belief, attributing an avoidance desire, and predicting 'opposite' behavior. Only one of these ways (avoidance desire) involves the overt use of negation. Although target-shifting may sometimes be involved in processing negation, we suspect that negation itself is not the crucial factor. Instead, we expect that the underlying mechanism involves an initial identification of one target answer, followed by a second step that disengages from that initial answer and shifts to another. For example, in avoidance desire, to determine in which box to place the fish, one attends to where the kitten is (believed to be), thus identifying what to avoid. Having identified what the character wants to avoid, inhibition is applied so that the index shifts away from the to-be-avoided target. Target-shifting by inhibition is what is critical, rather than negation or even desire itself. It should therefore be possible to generate target-shifting by inhibition by means other than avoidance desire.

In the next experiment, we test whether making judgements about a character's pretending can also recruit inhibitory target-shifting. We give subjects a scenario in which characters pretend that an object is in the opposite container to the one it is really in. We expect that an 'opposite pretend' scenario will involve target-shifting and, when combined with false-belief, will produce double inhibition and a pattern of performance similar to that found with avoidance desire and 'opposite behavior' in 4-year-old false-belief passers.

## 3. Experiment 2

### 3.1. Methods

#### 3.1.1. Subjects

Sixty-three children were seen. To be included, children had to pass a standard false belief screening task, as in the previous experiment. Nine children failed the screening task and were not tested further. Children were also rejected if they failed any of the control questions for a second time, following a repetition of the story up to that point. Seven additional children were rejected for that reason. The remaining 47 subjects (25 girls) were aged between 4 years 1 month and 5 years 0 months (mean age = 4 years 6 months, $SD = 2.9$ months), recruited from New Jersey preschools, with approximately equal numbers of girls and boys, randomly assigned to conditions with 20 children in the True Belief condition and 27 in the false-belief condition. Children were diverse in terms of SES and ethnicity, reflecting central New Jersey. English was the main language spoken at home in all cases.

#### 3.1.2. Materials

Props used were similar to those in the previous experiment: A male and a female doll, two differently colored boxes, and a toy banana.

#### 3.1.3. Design and procedure

Children were assigned to different groups in a between-subjects design. Both groups were told stories in which two protagonists, Mary and John, play a special "opposite" pretend game, whereby if there is a toy in box A then they pretend it is in box B, and vice versa. One group (True Belief) heard a version in which Mary places the toy in box A and later, while John watches, switches it to B. Children were asked two control questions, Memory and Reality. Subjects were required to answer these correctly (see above). Then Mary says, "John, go get the pretend toy." Children were then asked the Know question, "Does John know that the real toy is in here?" Again children were required to pass this question. Finally, subjects were asked the Prediction question, "Where will John look for the pretend toy?"

Subjects in the false-belief group heard a version in which Mary places the toy in box A and then John and Mary go home for dinner. While they are away another character switches the location of the toy to box B. Children were asked the Memory and Reality questions and required to answer correctly (see above). Then Mary and John return from dinner and Mary says to John, "John, go get the pretend toy." Children were asked the Think question, "Where does John think the real toy is?" and were required to pass this question. Finally, children were asked the Prediction question, "Where will John look for the pretend toy?"

### 3.2. Results

All children included in the analysis passed the control questions, and, as appropriate, the Know or Think questions. More children (70%) passed the True-Belief with Opposite-Pretend task than passed (41%) the False-Belief with Opposite-Pretend task (Upton's $\chi^2 = 3.87$, $p = .025$, one-tailed).

### 3.2.1. Discussion

All included subjects passed a standard false belief screen. Yet less than half were able to pass a false-belief task that included 'opposite pretend.' This produced an effect similar to that produced by combining false-belief with an avoidance desire (Experiment 1 (Sick Kitten condition); Cassidy, 1998; Leslie & Polizzi, 1998) and with opposite behavior (Experiment 1 (Mixed Up Man condition); Leslie & Polizzi, 1998). All the children in the False-Belief with Opposite-Pretend task passed the Think question in that task (and also the screening task). On this basis, for these children, an estimated failure rate for false belief is close to zero. An estimate of the combined failure rate for opposite-pretend with false-belief is therefore close to the failure rate of opposite-pretend-with-true-belief (30%). Yet the task was significantly harder (59% failure), suggesting the two sets of task demands interact.

In terms of target-shifting, we suggest that to predict where an opposite pretender will go, one first identifies where the real object is (believed to be). Having identified what to avoid, one then shifts to the other location. In a false-belief context, two target-shifts must be combined in order to select the final target. According to our models of the selection process, inhibition is required to produce a target-shift, and combining two target-shifts requires a higher level of inhibitory control than producing a single target-shift. If the combination is done in parallel (Model 1), then one inhibition must inhibit the other. If the combination is done serially (Model 2), then the final target selected is one that was previously inhibited.

### 3.2.2. Target-shifting and desire

According to our models, target-shifting is the common factor uniting false-belief, avoidance desire, opposite behavior, and opposite pretend. Our theory regarding target-shifting for desire is different from that for false-belief. False-belief requires a target-shift to overcome the default true-belief. We do not hold, however, that approach desire is a default that an avoidance desire must overcome. As Leslie and Polizzi (1998, p. 248) point out, the desire not to burn one's fingers is a perfectly ordinary desire. The reason that avoidance may require target-shifting is that often it is necessary to identify a target precisely in order to mark it as the *thing to-be-avoided.* Then to predict a character's avoiding action, that target must be inhibited so that the alternative target is selected. This account extends straightforwardly to opposite behavior and opposite pretend.

To test this account of avoidance desire, we can use a story in which the protagonist's desire is specified negatively, as being for "not X," but which does not require a target-shift at the time the Prediction question is answered. Accordingly, children should find such a task easy.

To test this hypothesis, we used a story about two dogs; one was all white, and the other was white with black spots. Children were told, "Sally wants to give a bone to the dog who does *not* have spots." We reasoned that, immediately upon hearing this specification, it is possible to identify the all-white dog as the target of Sally's desire.

That very dog can then be tracked as the desire target, without shifts, through the rest of the story. It seems unnatural to identify the dog-that-does-not-have-spots by first locating the dog that *does*, then selecting the other. If any desire target-shift occurs at all, it should be well before the critical moment when the Prediction question is being answered.

By contrast, in the sick kitten story, Sally's desire is for the *location* that does not currently contain the kitten and at some time the kitten occupies both. Intuitively, the kitten, rather than the locations, will be the center of the child's attention. The child will track the kitten as it moves location, rather than tracking kitten-less locations. In the Spotty Dog story, if the spots could somehow jump from one dog to the other, then perhaps target-shifting might be required. Without that, "not having spots" immediately and permanently identifies the all-white dog as the target. We therefore expected Sally's desire, "to give a bone to the dog that does *not* have spots," despite its complexity, to be a non-target-shifting desire. The Spotty Dog scenario when combined with false-belief should be a single and not a double inhibition scenario.

We tested 4-year-old standard false belief passers with two false-belief tasks with negatively specified desires. One task combined false-belief with a target-shifting desire and one combined it with a negatively specified but non target-shifting desire. We expected that the target-shift desire task would be harder for 4-year-olds than the non-shifting desire task. We expected the latter task to be equivalent to a standard task and therefore easy for these subjects.

## 4. Experiment 3

We compared a task hypothesized to be a single inhibition task with a task hypothesized to be a double inhibition task. To ensure that subjects were capable of making the false-belief inhibition, we required all to pass the Think question.

### *4.1. Methods*

#### *4.1.1. Subjects*

Twenty-eight children were seen. Only subjects who passed a standard false-belief screening task were included; 9 subjects failed the screening task and were not tested further. A further 3 subjects failed a Think question in one of the experimental tasks and data from these subjects were also excluded. The remaining subjects were 16 children (10 girls) aged between 4 years 2 months and 5 years 6 months (mean = 4 years 11 months, $SD = 4.6$ months). Testing took place in quiet areas of local schools; SES and ethnicity reflected central New Jersey diversity.

#### *4.1.2. Materials*

Story presentation was aided by props. These included two 3-dimensional Styrofoam model rooms, one for each of the two tasks. There were also two model doghouses and two boxes differing in color, assigned to one of the model rooms, respectively. In addition, there were various small dolls, toy dogs, and other props. The false-belief screening task used the same props as Experiment 1.

#### 4.1.3. Design and procedure

Following the screening task, each subject was given two tasks with order counterbalanced across subjects. In the non-target shifting task, a boy was described as wanting to give a bone to a dog that does *not* have spots. There were two doghouses, one containing a spotted dog and one containing an all white dog with no spots. The boy then went away to get the bone and while he was away the dogs switched places. For the target-shifting task, the same Sick Kitten (false-belief) story as Experiment 1, in which Sally does not want to give a fish to the sick kitten, was used.

For both tasks, subjects were asked two control questions: Memory question: "In the beginning, where was the [dog with no spots/sick kitten]?"; Reality question: "And where is the [dog that does not have spots/sick kitten] now?" Children were then asked a Think question: "Where does [protagonist] think the [dog that does not have spots/sick kitten] is now?". Finally, subjects were asked the Prediction question: "Which [doghouse/box] will [protagonist] go to with the [bone/fish]?"

Subjects who failed a control question had the story repeated up to that point. If they had failed the control question a second time, they would have been rejected from the study, but no child did so. Data from subjects who failed the Think question were not included in further analyses because we wanted to study only children who successfully attributed a false-belief. Three subjects were rejected for this reason.

### 4.2. Results

All 16 subjects gave the correct answer to the Prediction question in the Spotty Dog story whereas only 8 (50%) of these children correctly answered Prediction in the Sick Kitten story (McNemar Binomial, $N = 8$, $x = 0$, $p = .004$, one-tailed).

#### 4.2.1. Discussion

As predicted, specifying the protagonist's desire as "to give a bone to the dog that does not have spots" did not produce measurable difficulty for 4-year-old children who can reliably pass a standard false belief task. Where the target of desire can be identified 'at once' from a negative specification and then tracked throughout the rest of the story, as in the case of the non-spotty dog, children perform as well as with the approach desire of the standard false belief task we used as a screen. However, if a negative specification of desire is plausibly processed so that a target-shift occurs at the critical point in processing—around the time of the belief target-shift—then an otherwise manageable false-belief task becomes difficult.

In combination with the results of Experiment 2, these results rule out a number of alternative explanations for the difficulty that children have with avoidance desire. One possibility is that specifying an avoidance desire scenario is too complex for children to understand. The avoidance desire, "Sally does not want to give the fish to the sick kitten,'' is more complex than the approach desire in a standard task, ''Sally wants the marble.'' Could this additional complexity account for the difficulty of avoidance false-belief for 4-year-olds? We think not, for three reasons. First, children have no difficulty with the same avoidance desire when combined with a true belief. Second, Experiments 1 and 2 show that avoidance desire is not the only way that target-shifting can be created. Children in the opposite behavior condition (Experiment 1; see also Leslie & Polizzi, 1998) and in the opposite pretend task (Experiment 2) do not have to process a complex, negatively specified desire, yet have difficulty with the tasks. Third, the present experiment shows that a desire that is equally complex to state, ''Sally wants to give a bone to a dog that does not have spots,'' does not produce difficulty when combined with a false-belief. The complexity of the desire statement does not explain these effects. Instead, we argue that the effects should be understood in terms of target-shifting.

The next question that arises is whether we can actually measure the relative difficulty of target-shifting and non-target-shifting desires in 'isolation.' For a population of 4-year-olds who succeed on standard false belief, both desire and belief target-shifting is well within their capabilities, if each occurs singly. The difficulty of target-shifting is measurable only when two target shifts interact in belief-desire reasoning. However, for 3-year-olds who fail standard false-belief, the load of a *single* desire target-shift may be measurable, without adding a belief target-shift. If so, a target-shift desire would be harder than a non-target-shift desire in the context of a true-belief task. According to both models, attributing a true-belief is a default operation that does not require a target-shift. Thus, in the next experiment, we test 3-year-olds who fail standard false belief with a true-belief version of the Spotty Dog story and with the true-belief version of the Sick Kitten story from Experiment 1. Because the only target-shift required is for the desire in the Sick Kitten story, we hypothesized that this story would be harder for these subjects.

## 5. Experiment 4

### 5.1. Methods

#### 5.1.1. Subjects

Forty-seven children were seen. To be included in this experiment, subjects had to be both 3 years of age and fail a standard false belief screening task. Nine subjects were excluded for passing the screening task. An additional 11 children were tested but excluded, 5 for failing to complete all three tasks and 6 for repeatedly failing control questions. The remaining subjects were 27 children (16 boys) aged between 3 years 2 months and 3 years 11 months (mean age = 3 years 7 months, $SD = 2.5$ months) who failed the test question on a standard false-belief task while passing its control questions. SES and ethnicity reflected central New Jersey diversity.

#### 5.1.2. Materials

Same as Experiment 3.

#### 5.1.3. Design and procedure

Design and procedure were the same as Experiment 3, except for the fact that subjects had to fail the screening task to be included and the two test stories were administered in true-belief form. In true-belief form, the protagonist in the Spotty Dog story remains watching while the spotty and non-spotty dogs switch doghouses, and then he goes to get the bone. Instead of the Think question appropriate to the false-belief form, a Know question was asked: "Does [protagonist] know the dog with no spots is in this house?". All children answered this correctly. The other test story used was the Sick Kitten (True Belief) story from Experiment 1.

As in Experiment 3, the Spotty Dog story presented an avoidance desire that we predicted would not require a target-shift. The Sick Kitten (True Belief) story also presented an avoidance desire but one that requires a target-shift. Both stories were presented in true-belief form to allow us to compare directly the relative difficulty of target-shift and non-shift desire stories with 3-year-olds. Following the screening task, each child was given both test stories with order counterbalanced.

### 5.2. Results

Most children (89%) passed the Spotty Dog (True Belief) story, whereas only 59% passed the Sick Kitten (True Belief) story. Three children failed Spotty Dog but passed Sick Kitten, and 11 children showed the opposite pattern (McNemar Binomial, $N = 14$, $x = 3$, $p = .029$, one-tailed).

#### 5.2.1. Discussion

Nearly 90% of 3-year-olds who failed the standard false-belief task performed well on a true-belief task, even when the non-target-shifting desire was complex and negatively specified. This is further evidence that desire complexity itself is not the critical factor in the double inhibition effect. Although a majority passed the target-shift desire task, significantly fewer did so than passed the non-shift task. Desire target-shifting by itself then produces measurable difficulty for false-belief failing 3-year-olds. Given that none of our 3-year-olds passed a standard false belief task, we can also conclude indirectly that the target-shift demanded by the desire attribution in the Sick Kitten story is easier to produce than the target-shift demanded by a false-belief. In terms of our models, *less* inhibition is required to produce a target-shift in avoidance desire than in false-belief. This makes sense if a desire target-shift does not have to overcome a default. Desire inhibition can be weaker than belief inhibition and still be effective. The limited inhibitory control available to most 3-year-olds is sufficient for avoidance desire problems yet insufficient for false-belief problems.

#### 5.2.2. Latent and manifest difficulty

Target-shifts from avoidance desire, opposite behavior, and opposite pretend all appear to interact with the target-shift required by false belief attribution, depressing performance in otherwise successful 4-year-olds. This suggests that false-belief tasks *remain* difficult for 4-year-olds, even after they reliably pass these tasks.

Let us call the average failure rate that a given task produces in a group of subjects a measure of its *manifest* difficulty for that group. If a group of subjects is selected for their ability to pass a given task, then for that group that task will have zero manifest difficulty. However, one subject may pass a given task with ease—that is, with resources to spare—while another subject may pass that task only "by the skin of his or her teeth." For example, a typical 10-year-old child will find a standard false belief task insultingly easy, while a child around four may just make it and no more. Although both subjects pass, their processing resources are different. Alternatively, a comparison may be made between two tasks relative to a single group of subjects. Both tasks may be passed by this group, but one task is passed easily, while the other task is passed "just and no more." Conclusion: the two tasks differ in their processing demands. As Simon (1956) first made clear, what we generally want to model is the *balance* between the processing demands of a task and the processing resources a subject has available to solve it.

Let us call the degree of difficulty that a given task *fails* to measure for a given subject, its *latent* difficulty. (This term is meant to suggest a conceptual analogy with Black's *latent heat*, the heat of a body that cannot be measured by a thermometer.) If

$$\text{latent difficulty} = \text{task demand} - \text{available resources},$$

then the latent difficulty of the standard false belief task is that balance of demand and resource that goes unmeasured when a child passes or fails. In the case of failing, demand exceeds resource and latent difficulty has a positive value whose magnitude measures how far the child is from passing. In the case of passing, resource exceeds demand and latent difficulty has a negative value, the absolute magnitude of which measures how far the child is from failing.

One way to think about the double inhibition task is that it reveals the latent difficulty in the standard task. If the task demand of false belief were reduced, sparing resources, then children's performance level on the double inhibition task might rise. A task manipulation that may reduce latent difficulty is the 'look first' question, to which we now turn.

#### *5.2.3. Seeking an answer to the 'look first' question*

When the Prediction question in a standard false belief task is changed minimally to contain the word *first*, "Where will Sally look *first* for her marble?" 3-year-old performance improves (Siegal & Beattie, 1991; Surian & Leslie, 1999; Wellman et al., 2001). Surian and Leslie (1999) found that 3-year-old children who fail the standard Think question are helped by a 'look first' Prediction question, whereas older children with autism are not. Control tasks show that the 'look first' question does not simply produce more "first location" responses regardless of belief attribution. "First location" responses are not forthcoming in true-belief versions of 'look first,' where the first location of the target is the wrong answer (Siegal & Beattie, 1991; Surian & Leslie, 1999). The effect of the 'look first' form of the question is therefore sensitive to epistemic status, revealing competence earlier than the standard question form.

Siegal and Beattie (1991) proposed that failure by 3-year-olds on standard false belief tasks was related to limitations in their pragmatic skills. In essence, 3-year-olds fail to 'get the point' that the experimenter is asking about belief-driven search rather than about successful search. They are helped by the 'look first' format because this makes the experimenter's intentions clearer (see also Clements & Perner, 1994 for a similar suggestion).

Surian and Leslie (1999) point out that, even if this idea is right, it leaves unexplained why the 3-year-old needs this help and the 4-year-old does not. What is it about the way false-belief tasks are processed that changes with development and gives rise to these changing needs? Surian and Leslie suggest a possible answer. The 'look first' question draws the child's attention to the location where the object was first and implicates further looks after Sally's first look in a failing location. The result is increased salience for the first location. Results from true belief control tasks show that this increased salience does not operate in a 'dumb' way, but is sensitive to Sally's epistemic status. The first location acquires increased salience *as the target of Sally's belief*. The increased salience for the false-belief brings it more into balance with the default true-belief, and, in turn, reduces the strength of the inhibition needed for its selection. This explains how the young child better grasps what the questioner is 'getting at' and also, as inhibitory resources increase, why the older child no longer needs this help.

Surian and Leslie's speculations connect the 'look first' effect with the ToMM–SP model. They suggest that 'look first' works by reducing demand for inhibitory control, the same resource for which avoidance-desire increases demand. What can we predict in regard to the 'look first' effect in 4-year-old (standard) passers? At first glance, the question seems senseless: if a child passes standard false belief *without help*, how could she be helped any further? But that is to think only about manifest difficulty—the difficulty measured by pass/fail rate. The 'look first' question might still reduce the *latent* difficulty of the task, even when the manifest difficulty is zero, driving down an already negative latent difficulty value.

We can test this by examining whether 'look first' helps 4-year-olds in 'double inhibition' tasks. What do our models predict? When we examined this question, it turned out that the two models gave different answers and the opportunity to distinguish them experimentally.

#### *5.2.4. Recalculation versus reuse*

Suppose we ask a 4-year-old an arithmetic question like '$2 + 2 = ?$' and suppose that the child succeeds in getting the answer. If we then immediately asked this child to add one to his or her answer, we would expect the child simply to *reuse* the first answer and make the easy calculation '$4 + 1$.' However, it is conceivable that the child might start all over again and *recalculate* the first answer by solving '$2 + 2 + 1$.' With the distinction between reuse and recalculation in mind let us consider our models.

In the avoidance false-belief task children are first asked a Think Question, at which time the belief content is identified. Having solved the belief problem, do children *reuse* or *recalculate* this answer when the Prediction question is asked? According to Model 1, belief and desire are identified in parallel. Because Model 1 assumes parallel identification, it is simply not able to reuse a previous identification of belief.

(And if it did, then it would be identifying belief and desire serially, making it Model 2, not Model 1.) To answer Prediction following Think, recalculation of belief is mandatory for Model 1. In contrast, with the serial operation of Model 2, reuse of a previous belief identification is perfectly possible.

Whether or not the 'look first' format question can improve performance on a double inhibition task hinges on the distinction between recalculation and reuse. If a previous belief answer (to Think) is reused in answering the 'look first' Prediction question, then the manipulation has no opportunity to help. Only if belief is calculated again, as in Model 1, will 'look first' have an opportunity to reduce latent difficulty. In sum: Because Model 1 requires recalculation, it predicts that 'look first' will help 4-year-old standard passers on the double inhibition task. Because Model 2 allows reuse of the Think answer, it does not predict that 'look first' will help.

In the next experiment, we test whether the 'look first' question format helps 4-year-olds who are already successful belief-desire reasoners. Model 1 says it will, Model 2 says it will not. A simple clarification account also predicts no help. Because subjects will be required to pass a standard false belief task and to answer correctly the (standard) Think question immediately before 'look first' is asked, they will have demonstrated an already clear grasp of the experimenter's intention to ask about false-belief. There is just no room for further clarification.

## 6. Experiment 5

### 6.1. Methods

#### 6.1.1. Subjects

Twenty-eight children were seen. Subjects had to pass a standard false belief screening task, including its control questions, and also pass the Think question in the main experimental task. Ten children failed the screening task and an additional 2 subjects failed the Think question. The remaining subjects were 16 children (10 girls) aged between 4 years 0 months and 5 years 0 months (mean age 4 years 7 months, $SD = 3.7$ months). Testing was done in quiet areas of local schools; SES and ethnicity reflected central New Jersey diversity.

#### 6.1.2. Materials

Story presentation was aided by props. These included a 3-dimensional Styrofoam model room, two boxes differing in color, a small doll, a toy cat and a toy fish. The screening task used the same props as in Experiment 1.

#### 6.1.3. Design and procedure

Children who passed the screening task were introduced to a new set of props and told the Sick Kitten (false-belief + avoidance-desire) story, as in Experiment 1. Subjects were then asked four questions: The Memory question: "In the beginning, where was the kitten?"; the Reality question: "Where is the kitten now?"; the Think question: "Where does Sally think the kitten is?"; and the 'look first' Prediction question: "Where is the first place that Sally will try to put the fish?".

### 6.2. Results

Thirteen out of 16 (81%) subjects correctly answered the 'look first' Prediction question. This proportion is significantly greater than would be expected by chance (Binomial test, $N = 16$, $x = 3$, $p = .011$, one-tailed). We also compared this performance with the findings from Experiment 3, Sick Kitten (avoidance-desire + false-belief) group, in which 8 out of 16 passed the regular format Prediction question. The results of this comparison show a significantly greater proportion passing 'look first' Prediction than regular Prediction (Upton's $\chi^2 = 3.35$, $p = .034$, one-tailed).

#### 6.2.1. Discussion

When the Prediction question used the 'look first' format, the proportion of 4-year-old standard passers who could also pass the false-belief + avoidance-desire task was high. In fact, performance was back to the level typical of a group of 4-year-olds on a standard (single shift) false-belief task. More importantly, significantly more children passed with this question format than in Experiment 3 with the regular format, despite the fact that Experiment 3 produced the best level of performance seen so far in both the present series and in previously published results.

Before we can reach any firm conclusions, the results of Experiment 5 must prove replicable. Furthermore, we acknowledge that the wording of the Prediction question differed in more than the word 'first.' We felt that asking "Where will Sally put the fish first?" sounded infelicitous because Sally would surely notice the kitten before actually placing the fish there. We therefore chose the wording "Where's the first place Sally will try to put the fish?" However, this introduces the verb "to try." Perhaps that phrase contributed to the difference we found between the 'look first' and regular formats for the Prediction question. We therefore conducted a new experiment in which we changed the wording of the regular Prediction question to "Where will Sally try to put the fish?" We then ran both regular and 'look first' question format conditions. We also duplicated these conditions using the Mixed-up Man story so that we could compare performance by question format with an opposite behavior target-shift. We also ran the true-belief versions for all of these conditions needed to control for low-level response strategies.

A final issue addressed in this study concerned the possibility that children might have difficulties in the target-shift tasks because they were failing to remember all of the relevant information at the time the action prediction question is asked. Some theorists working in the tradition of executive function limitations on children's belief-desire reasoning have stressed working memory rather than inhibitory demands in the standard false belief task (Gordon & Olson, 1998; Davis & Pratt, 1995). It might be argued that the target-shift problems create additional memory demands because the belief and action prediction responses mismatch. Incorrect responses to the action prediction question might occur if children failed to recall that the desire was to avoid. Accordingly, we included a reminder of Sally's avoidance-desire or the ‘oppositeness’ of the Mixed-up Man’s behavior, immediately prior to the test question, after the children had answered the belief/knowledge control questions.

## 7. Experiment 6

### 7.1. Method

#### 7.1.1. Subjects

Sixty-seven children were seen. To be included, subjects were required to pass a standard false belief task; 3 children were excluded for failing the screen and were not tested further. In addition, one child passed the screen but then refused to answer any further questions and was therefore excluded. The remaining subjects were 63 children (32 girls) aged between 4 years and 0 months and 5 years 8 months (mean age = 4 years 9 months, *SD* = 5 months), randomly assigned to one of four groups. Children were tested in quiet areas of local schools; SES was diverse and ethnicity predominantly Caucasian reflecting the population of Essex, England.

#### 7.1.2. Materials

Materials were the same as in Experiment 1.

#### 7.1.3. Design and procedure

Experiment 1 was repeated with some modifications. The first modification was to include a ‘look first’ condition together with a regular question condition. The second modification was to change the wording of the Prediction questions. In the ‘look first’ condition, the Prediction question was, “Where’s the first place Sally will try to put the fish?”, while in the regular condition the Prediction question was, “Where will Sally try to put the fish?”. The final change was that in all conditions, before asking the Prediction question, we gave the child a reminder of the protagonist’s desire (Avoidance Desire tasks) or of the protagonist’s disposition (Opposite Behavior tasks). This last change should produce better performance on the false-belief stories, if remembering the nature of the desire or disposition is a source of difficulty.

Each of the two tasks, Avoidance Desire and Opposite Behavior, was presented in both true- and false-belief versions, which, together with regular versus ‘look first’ Prediction questions, yielded a total of eight conditions. Each subject was randomly assigned to two of the eight conditions with the constraint that no child received both true- and false-belief versions of the same story, and no child was assigned to both a regular and a ‘look first’ condition. Sixteen children participated in each of the two regular Prediction conditions, 17 children participated in ‘look first’ Avoidance Desire true-belief/Opposite Behavior false-belief condition, and 14 children participated in the ‘look first’ Avoidance Desire false- belief/Opposite Behavior true-belief condition.

Subjects who failed Memory or Reality (control) questions were corrected the first time, then the story was retold up to that point and the control question asked again.

A second failure would have meant rejection but in fact no child failed a control a second time. To maintain pragmatic naturalness, subjects were asked a Know question in true-belief conditions and a Think question in false-belief conditions. As in Experiment 1, we treated the Know and Think questions as control questions and adopted the same procedure as for the Memory and Reality questions. Subjects who failed were corrected, recycled through the story, and asked the Know or Think question again. No subject failed these questions a second time. Following the reminder, all subjects were asked a Prediction question.

In true-belief conditions, passing requires indicating the location opposite to where the object is in reality. In the false-belief conditions, passing requires indicating the box where the object actually is. Better performance was predicted in true-belief than in false-belief conditions. Following Surian and Leslie (1999), it was predicted that the 'look first' question format would produce better performance over the regular format Prediction question in the false-belief conditions, while performance in the true-belief conditions would not be affected by the form of the question.

### 7.2. Results

In the regular format Prediction conditions, 16 of 16 subjects (100%) answered correctly in the Avoidance Desire True-belief task, while only 4 of 16 subjects (25%) answered correctly in the Avoidance Desire False-belief task (Upton's $\chi^2 = 18.6$, $p < .001$, one-tailed); 15 of 16 subjects (94%) were correct in the Opposite Behavior true-belief condition, while 7 of 16 (44%) passed in the false-belief condition (Upton's $\chi^2 = 9.02$, $p = .001$, one-tailed).

In the 'look first' conditions, 16 of 17 subjects (94%) answered correctly in the Avoidance-Desire True-belief task and 10 of 14 (71%) were correct in the false-belief version (Fisher's Exact, $p > .1$, one-tailed); 12 of 14 subjects (86%) were correct in the Opposite-Behavior True-belief condition and 14 of 17 (82%) passed in False-belief (Fisher's Exact, $p > .2$).

No subject failed the Know or Think questions in the regular Prediction conditions. In the 'look first' conditions, no subject failed a Know question, one subject failed a Think question on the first round in the Avoidance Desire task, and two failed Think on the first round in the Opposite Behavior task. Eliminating those subjects did not change the results: Avoidance Desire True- versus False-belief, 94% versus 69% (Fisher Exact, $p = .09$, one-tailed); Opposite Behavior True- versus False-belief, 86% versus 80% (Fisher Exact, $p > .2$, one-tailed).

The comparisons of key interest focus on the format of the Prediction question. Comparing frequencies of subjects passing/failing regular format questions versus 'look first' format questions in true-belief tasks showed no significant difference (Fisher's Exact, $p > .3$). In contrast, question format made a substantial difference in false-belief tasks (Fig. 4). The 'look first' form of the Prediction question produced more correct responses than the standard format in both Avoidance Desire stories (Upton's $\chi^2 = 6.25$, $p = .006$, one-tailed) and Opposite Behavior stories (Upton's $\chi^2 = 5.15$, $p = .012$, one-tailed). Eliminating the first round Think question failers

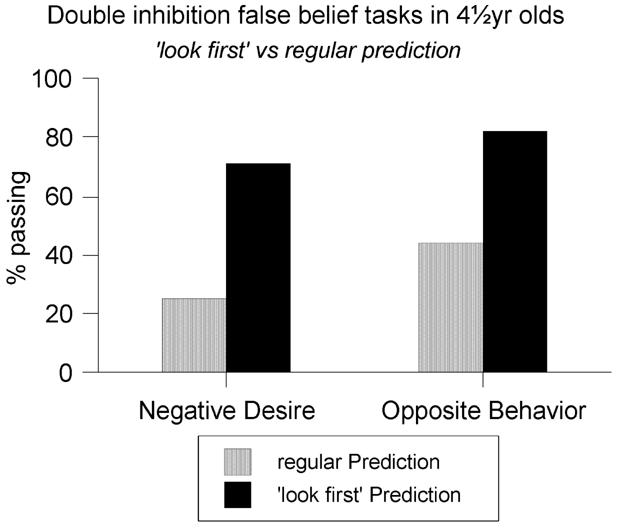


Fig. 4. Experiment 6: Standard false belief task passers are helped by a 'look first' question format in both avoidance false-belief and opposite-behavior false-belief scenarios.

does not change this result: Avoidance Desire, Upton's $\chi^2 = 5.48$ ($p = .01$, one-tailed) and Opposite Behavior, Upton's $\chi^2 = 4.15$ ($p = .021$, one-tailed).

### 7.3. Discussion

The 'look first' format for the Prediction question helped 4-year-old standard false belief task passers to also pass double inhibition false belief. We modified the wording of the Prediction questions such that in the regular format we asked, "Where will Sally try to put the fish?", and in the 'look first' format we asked "Where's the first place Sally will try to put the fish?." Changing the wording of the regular format question by inserting the "try to" phrase did not change performance; nor did inserting a reminder about the protagonist's desire (or behavior disposition). In line with previous findings, performance on double inhibition remained poor at 25% (avoidance desire) and 44% (opposite behavior). However, in the 'look first' condition, adding a "first place" phrase to this question produced a substantial improvement to 71 and 82% passing, respectively. Finally, the 'look first' effect was specific to false-belief; performance on true-belief scenarios remained high, unaffected by question format. The present results thus confirm and extend the findings from Experiment 5.

The finding that a reminder of the character's avoidance desire or opposite behavior had no discernable effect on children's performance suggests that the difficulty of target-shift tasks is not simply the result of forgetting (Davis & Pratt, 1995; Gordon & Olson, 1998). If children's failure in our earlier experiments resulted from failure to recollect the avoidance desire, we would expect to see improved performance in this study, where this information was re-presented immediately prior to the Prediction question, and after the children answered the Think question. It seems unlikely that working memory problems alone can account for the range of data presented here, though it remains possible that working memory might play a role *together with* inhibitory demands (Carlson, Moses, & Breton, 2002).

Once again we found a wide divergence between performance on the (standard) Think question and the regular format Prediction question—100% versus 25% passing (avoidance desire) and 94% versus 44% (opposite behavior). Such divergence, completely unknown in standard tasks, appears to be characteristic of the double inhibition task. When the Prediction question is in 'look first' format, however, divergence is greatly reduced or eliminated.

The finding that the 'look first' format in the double inhibition task impacts false-belief but not true belief performance echoes the findings of Surian and Leslie (1999) with 3-year-olds on standard false belief. Surian and Leslie (and before them Siegal & Beattie, 1991) included a true belief task to control for the possibility that a 'look first' question might simply bias children to point to the first location of the bait, generating apparently correct answers but without calculating belief. In a true belief task, pointing to the first location of the bait is a wrong response. A low-level response strategy for 'look first' that improves performance on false-belief will depress performance on true belief. Good performance on *both* tasks is required to establish the effect. In the double inhibition task, the pattern of correct responding is reversed relative to the standard task: the correct answer for true belief is the first (now empty) location, while for false-belief it is the second (now full) location. Despite this response reversal, the present results mirror previous findings from 3-year-olds that 'look first' improves performance on false belief while leaving intact good performance on true belief. Our results, therefore, underline the conclusion that 'look first' eases the selection of the false-belief content.

#### *7.3.1. 'Look first' and inhibition*

Most importantly, the results of Experiments 5 and 6 together show that the effect of 'look first' is not limited to helping the 3-year-old who otherwise fails standard false belief. The addition of the same small phrase also transforms the performance of the 4-year-old who passes standard false belief but who would otherwise fail double inhibition false-belief. However, in the case of our 4-year-olds, it is not possible to put forward an explanation of the 'look first' effect in terms of clarification. It is already quite clear to them that the experimenter intends to ask about Sally's false-belief: All our subjects passed a standard false belief task, and all also correctly reported Sally's belief just prior to answering Prediction. Therefore, these children require neither clarification of experimenter's intent (Siegal & Beattie, 1991) nor clarification of the temporal structure of the story (Wellman et al., 2001). For simple clarification accounts of how 'look first' helps 3-year-olds, it is wholly surprising that 'look first' should also help 4-year-olds who already understand false belief without 'clarification.' We conclude that 'look first' must be helping them some other way.

Instead of having two completely different accounts of the effect of 'look first,' one for 3-year-olds and another for 4-year-olds, it is better to look for a single account that covers all the data. Surian and Leslie (1999) suggested that the 'look first' format acts by reducing the inhibitory demands of the false-belief task. Although Surian and

Leslie made this suggestion for 3-year-old standard task failers, inhibition reduction will also explain why the 'look first' format helps 4-year-old standard task passers with a double inhibition task.

Earlier we drew a distinction between manifest and latent difficulty. This highlights the relationship between the demands on problem solving resources a task makes and the resources a child has available. Although a standard false belief task may be reliably passed by a 4-year-old, the task retains a certain degree of latent difficulty that may be made manifest again by avoidance-desire false-belief. The manifest difficulty of the standard task for the 3-year-old can be reduced by the 'look first' question format. Likewise, the manifest difficulty of the avoidance-desire version for the 4-year-old can be reduced the same way. Put simply, and somewhat tongue-in-cheek, by manipulating the inhibitory demands of false-belief tasks, we can turn 3-year-olds into 4-year-olds, 4-year-olds into 3-year-olds, and then turn 4-year-olds who would have been turned into 3-year-olds back into 4-year-olds!

The assumption at the heart of the SP model is that successful false belief calculation involves at least two candidate belief contents, one of which is the true-belief content and the other of which will be a plausible but alternative false-content (typically corresponding to something relevant that the protagonist saw or heard). The true-content starts out being more salient than the false-content but is subject to review. We surmise that the 'look first' question draws attention to the false-content and renders it more salient that it otherwise would be. The question format thereby tends to reduce the salience *differential* between true- and false-contents. The reduced differential in turn requires less inhibition to reverse its direction, thus easing the selection of the false-content and thereby improving performance.

As Surian and Leslie pointed out, the inhibition reduction account of 'look first' does not so much contradict Siegal and Beattie's account as provide a mechanism for how clarification might work. 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 in order to pass standard false belief Prediction. To this we can now add that the same mechanism also accounts for why 4-year-olds still can benefit from 'look first' when faced with a more difficult double inhibition task.

#### *7.3.2. Lock and key*

Siegal and Beattie (1991) discovered that a minimal change to the standard false belief task—the addition of the word "first"—has a disproportionate effect on performance. Cassidy (1998) discovered that another minor change—the addition of the word "not" to the desire specification—can also have a disproportionate effect on children's false-belief reasoning.

Lets one imagine that just about *any* change made to the wording of these tasks will have large effects on young children, we remind the reader that that has *not* been the experience of researchers over the last 20 years. Investigators from Wimmer and Perner (1983), through Gopnik (1993), to Wellman et al. (2001), have rightly stressed how immune are the effects in false-belief tasks to even quite large task changes. For example, it makes no difference whether one asks the child about someone else's belief in an unexpected location task, like "Sally and Ann," or about the child's own belief in a deceptive appearances task, like "Smarties" (Perner, Leekam, & Wimmer, 1987). Indeed, "the extent and variety of task materials, manipulations, questions, and controls that investigators have used" to such regular effect persuaded Astington and Gopnik (1991) that "3-year-olds truly have a conceptual deficit" (p. 11). However, despite the agreed robustness of the findings, we believe their conclusion was premature.

From our perspective, the child employs heuristic reasoning. As such, the child's cognitive mechanisms are like a lock and task structure like a key. There are relatively few changes one can make to a key that will systematically switch it from being ineffective (cannot open lock) to effective, from effective to ineffective, and back again to effective. Likewise with changes in task structure that enable/disable a child's problem solving. But systematic changes that do work like this are golden because their careful study can lead to an understanding of, in Hume's famous phrase, the "secret springs and mechanisms" of the mind.

Finally, we note that our findings on 'look first' were predicted by Model 1 but not by Model 2. We discuss this point more fully below.

## 8. General discussion

Four-year-olds have difficulty with false-belief problems involving a protagonist's desire to avoid rather than to approach the target. This difficulty cannot be attributed to a conceptual deficit or to an inadequate theory of belief. The ToMM–SP framework provides an explanation in terms of performance factors: whereas answering the Think question requires control of only a single inhibition, correctly answering Prediction requires control of a double inhibition. It has long been known that *some* performance factors must enter 'theory of mind' reasoning or why else would there be a delay between passing first-order and second-order false-belief problems (Perner & Wimmer, 1985)? However, within the ToMM–SP framework we have developed specific hypotheses about the nature of the performance factors involved. These hypotheses extend beyond avoidance tasks at 4 years to account for failure on standard tasks at 3 years. They can also explain the nature of the 'help' offered by certain task manipulations at both three years (standard tasks) and at four years (avoidance tasks). Because second-order false-belief tasks also involve a kind of double inhibition perhaps the same hypotheses may explain their difficulty too.

More specifically, the six experiments we report allow us to advance a number of conclusions regarding early belief-desire reasoning. We list these below.

(1) We replicated the results of two previous studies (Cassidy, 1998; Leslie & Polizzi, 1998) with avoidance-desire + false-belief tasks. Such tasks are substantially more difficult than the standard approach-desire + false-belief tasks usually used to assess early 'theory of mind' performance. Only a minority of 4-year-old children who can pass a standard false belief task can also pass Prediction in the avoidance task: 12% in Experiment 1, 50% in Experiment 3, and 25% in Experiment 6. The effect appears to be reliable and quite large. A similar picture

emerged when we compared children's performance on the Think and Prediction questions within the avoidance-desire + false-belief task, with almost all subjects who failed Prediction passing the Think question and none showing the opposite pattern. Finally, this whole pattern of findings was duplicated in tasks in which a protagonist was described as having approach desires but who habitually performed the 'opposite' action to that which would satisfy the desire.

(2) The above effects remain after the child has been reminded of the protagonist's desire or opposite-behavior disposition immediately prior to questioning (Experiment 6 standard Prediction conditions).

(3) Negation in a desire specification is not sufficient for poor performance in the avoidance-desire + false-belief task (Experiment 3). Instead, the key appears to be a desire specification that calls for a 'target-shift' at the point in processing when a behavior prediction is selected. If the subject can identify the target of desire early in the task and then track that target through the remainder of the story (as in the 'Spotty Dog' false-belief story), then there is no need to shift desire targets at the critical point in processing. In this case, the story is equivalent in difficulty to a standard false belief task because only a single inhibition (belief) is required. Negation is also not necessary for poor performance: Target shifts can be produced in a number of different ways, including false-belief, opposite behavior, and opposite pretend (Experiments 1, 2, 3, 4, 5, and 6).

(4) For 3-year-olds, an avoidance-desire that produces a target-shift (Sick Kitten true-belief story) is significantly harder than a similar task (Spotty Dog true-belief story) that does not require a desire target-shift (Experiment 4). All of the children who passed the Sick Kitten true-belief task failed a standard false belief task. This shows that a desire target-shift is easier to produce than a belief target-shift. A weaker inhibition may shift a desire target while a stronger inhibition is required to shift a belief target. Perhaps this is because a belief, but not a desire, inhibition must overcome a default.

(5) The 'look first' question format helps 4-year-olds pass double inhibition tasks, demonstrating that it continues to help children even after they pass the standard task. Accounts of how the 'look first' question helps must address this new finding. Simple question clarification cannot account for its helping the child who already correctly attributes false-belief.

(6) The 'look first' question helps children pass both avoidance-desire and opposite-behavior false-belief tasks without hindering them from passing the corresponding true belief tasks. Since the correct answers in true- and false-belief tasks are opposite locations, this result rules out low-level responding, such as, the first location, a location where search will fail, a location other than the actual location of the target, and strategies that affect desire attribution or behavior prediction *directly* (for example, desire or go-to the opposite location). If any of these strategies were employed in response to 'look first,' they would result in good performance on false-belief and poor performance on true belief, not good performance on both. Therefore, the psychological site of impact for 'look first' must be the calculation of belief content.

(7) Finally, because the 'look first' question impacts the calculation of belief contents, in order for it to have this effect, the child must actually calculate belief in response to the 'look first' question. And here, our findings show that the effect occurs even just after the Think question has been answered correctly. If the result of this previous calculation of belief (in response to Think) was simply carried over and taken as the starting point for answering Prediction, then 'look first' would not get an opportunity to have that impact. Any costs of calculating false-belief would have already been paid and could no longer be reduced. Following this line of reasoning, we conclude that *re*calculation of belief takes place in response to the Prediction question.

### 8.1. Recalculation of belief. But why?

From a pre-theoretical point of view, recalculation seems unexpected, even odd. Commonsense would suggest that if you have struggled to answer a difficult question and then are asked an easy question that follows on from that answer, then you should simply start with that answer. To return to our example of the 4-year-old who succeeds in getting the answer to the difficult arithmetic question '$2 + 2 = ?$' and who is then asked to add one to that: We would hope the child would undertake the easy calculation '$4 + 1$.' We would not expect the child to start all over again and try to calculate '$2 + 2 + 1$.' In the case of false belief, commonsense suggests that, having grasped that Sally thinks that the cat is in the basket, the child should simply figure that the basket is what Sally wants to avoid. Done that way, one piece at a time, the task should be easy for a 4-year-old. The fact that it is not easy suggests that Prediction is accomplished by calculating belief and desire in the same process. And if belief has already been calculated for Think, it must be recalculated along with desire for Prediction.

Exactly one of our models predicts the recalculation. Recall that the difference between the models is whether belief and desire are identified in parallel (Model 1) or serially (Model 2). Because Model 1 assumes parallel identification, it cannot reuse a previous identification of belief. If it did, then it would simply be identifying belief and desire serially, and this would make it Model 2, not Model 1. Therefore, according to Model 1, recalculation of belief is mandatory. For that reason, the findings on 'look first,' showing that recalculation does in fact take place, support Model 1.

What is the situation with Model 2? Model 2 can allow reuse of a previously made identification of belief (in response to Think) by simply adding a desire index in response to Prediction. To see this, the reader may refer to the last panel of Fig. 4; this panel simply combines the target shift of the second panel with the target shift of the third panel. Model 2 therefore does not *require* recalculation of belief. To be sure, Model 2 could be made consistent with recalculation—by an additional assumption that, for example, a previous identification cannot be remembered. However, Model 1 actually *predicts* these findings.

A second piece of evidence also points to Model 1. Recall that a large number of studies show that for *standard* tasks there is no measurable difference between performance on the Think and Prediction questions, despite the fact that Prediction requires desire and action to be considered in addition to belief. Model 1 has a ready explanation for this, namely, that an approach desire requires no further work after calculating the false-belief. Again Model 2 can made consistent with this by supposing that the allocation of a desire index has no cost. But given this is not required by the model, Model 2 does not predict this fact. Indeed, only Model 1 predicts: (a) no difference between Think and Prediction questions in standard tasks, (b) greater difficulty for Prediction over Think in avoidance false-belief tasks, and (c) mandatory recalculation. Therefore, subject to further research, the data favor Model 1.

### *8.2. Alternative accounts: The theory–theory*

Current 'theory–theories' offer few hints as to how one might explain the present findings. The 4-year-olds who pass approach-desire false-belief but fail avoidance-desire false-belief clearly employ the concept BELIEF, without a conceptual or theory deficit. It is hard to see why understanding a wish to avoid rather than to approach something would require a more advanced theory. Indeed, Wellman's account of the child's 'theory of desire' explicitly includes avoidance in the child's early desire-theory (Bartsch & Wellman, 1995, p. 12). It is harder still to see why a new theory of belief would be necessary. Our findings point instead to performance factors.

It is common ground between most researchers that any serious account must include both the nature and origins of metarepresentation (competence) and the processing (performance) systems employing it (see e.g., Wellman et al., 2001, pp. 656–657). What controversy has centered around is how to understand the changes in children's response to belief-desire problems observed between the third and fifth birthdays. For one group of researchers, the observed changes are evidence of change in conceptual competence (e.g, Perner, 1991; Wellman et al., 2001), with at most a minor role for performance change. Crucially, the hypothesized conceptual change over this period is assumed to show that the concept of belief is a *complex construction learned through experience*.

For another group of researchers, the observed changes in responding between the third and fifth birthdays provide evidence only for performance change (e.g., Bloom & German, 2000; Leslie, 1994a, 2000a, 2000b; Scholl & Leslie, 2001). Given that Wellman et al., 2001 allow at least *some* role for performance change over this period, both groups agree the choice is between a conceptual-change-*plus*-performance-change account and a performance-change-*only* account.

To date task manipulations have improved 3-year-old performance but only to levels short of ceiling. Wellman et al. interpret this as providing support for conceptual change. However, as Scholl and Leslie (2001) point out, the data are merely consistent with conceptual change but do not support it. There is no reason to suppose that only a single performance factor is at work in these (or any other) complex tasks, nor that task manipulations to date must have exhausted those factors, nor that one can ever entirely 'remove' performance factors. This is because performance factors are *not* task 'artifacts,' as Wellman et al. (2001) call them, that need to be removed by appropriate controls, but rather properties of the underlying processing system that need to be understood.

In our view, the observed changes in responding from 3 years onward are straightforwardly explained by a performance-change-only account. Conceptual change after 3 years remains an empirical possibility, but, *pace* Wellman et al. (2001), only a possibility. Although our present experiments address this question only indirectly, we have reasons which we review briefly below for supposing that conceptual change does not take place in this period. But first we consider alternative performance accounts.

### 8.3. Alternative performance accounts

How do other currently available performance accounts fare with these data? Most current alternatives are designed to account for 3-year-old failure and say little about how 4-year-olds actually achieve success, beyond implying that whatever was 'broken' is now 'fixed.' Such accounts will generally not explain why 4-year-old passers should find avoidance-desire false-belief hard or why 'look first' should help them. For example, in syntax-based accounts of false belief reasoning (DeVilliers & DeVilliers, 2000), it is hard to see why, when the protagonists' desire is to avoid the target, syntax-based success should turn into syntax-based failure. Perhaps, the newly successfully child *just* manages to parse correctly and any extra burden at all—a 'last straw'—will sink her. But then why does a question with a yet more complex construction—"Where's the first place Sally will try to put the fish?"—make the task easier than the syntactically simpler question, "Where will Sally put the fish?" Again, two or more unrelated accounts might work. But a single unified account is better, and Hale and Tager-Flusberg (2003) have recently shown how to integrate the role of language into the ToMM–SP framework.

Similar remarks apply to the proposal that 3-year-olds fail standard tasks because they cannot reason with counterfactual propositions (Riggs & Mitchell, 2000). When this deficit is 'fixed,' the older child succeeds. However, because avoidance-desire adds no additional *counterfactual* material to the task, it is unclear how this account would predict either its difficulty or the easing role of the 'look first' format. One might appeal to a 'last straw,' but again it is hard to see why a desire to avoid should be a 'last straw' or why 'look first' should remove the 'straw' again. On the other hand, counterfactual reasoning may also place demands on inhibitory executive processes suggesting a possible integration. (see German & Nichols, 2003).

Some current accounts have features in common with the selection processing framework. Mitchell (1994) has suggested that young children fail false-belief tasks because they have a general 'reality bias' which draws them to indicate where the bait really is rather than indicate where the protagonist thinks it is (see also Russell, Mauthner, Sharpe, & Tidswell, 1991). A 'reality bias' and a 'true-belief bias' are inevitably closely related because, for the belief attributer, 'reality' and what is 'true' are the same. However, the biases are very different psychologically. Attributing a true belief is part of 'theory of mind'; responding to reality is not. A non-mentalizing bias to respond to reality will not account for 4-year-old failure on avoidance-desire. When a child fails an avoidance-desire false-belief task, she fails to point at where the bait really is, the response predicted by a 'reality bias.'

Other related accounts framed in terms of executive function (EF) failures (e.g., Russell et al., 1991) highlight variously the roles of inhibitory control (Carlson, Moses, & Hix, 1998), working memory or 'holding in mind' (Gordon & Olson, 1998; Davis & Pratt, 1995), or a combination of both (Carlson et al., 2002). The ToMM–SP model is generally compatible with these ideas.

Fodor's (1992) model and ToMM–SP share many background assumptions. However, the MO that Fodor proposed had 3-year-olds routinely ignore belief, whereas ToMM–SP has both three- and 4-year-olds routinely calculate belief, with higher costs for calculating false-belief.

### *8.4. The role of executive functions in 'theory of mind'*

Moses (2001) usefully distinguishes between two roles that executive functions (EF) may play in "theory of mind" development. The first is *expression*: the BELIEF concept already exists in young children but struggles to find expression in performance because of limitations on EF. This is roughly the kind of role envisioned in the ToMM–SP framework.

The second possible role is *emergence*: the BELIEF concept is constructed by the child and EF plays a key role in the construction process. This is the role favored by Moses and colleagues (Carlson et al., 1998, 2002; Carlson & Moses, 2001; see also Perner & Lang, 1999; Russell, Saltmarsh, & Hill, 1999; for a working memory emergence account, see Davis & Pratt, 1995; Gordon & Olson, 1998). Our present findings only indirectly address the expression/emergence issue. However, we will make two points.

First, if the concept BELIEF is constructed—constructed out of other concepts—we need to know *which* concepts these are, and then how *those* concepts were acquired. The only specific proposal is that of Perner (1995), who argues that BELIEF is constructed out of the concepts, SEMANTICAL, EVALUATE, MENTAL, REPRESENTATION, EXPRESSES, and PROPOSITION. Note that each one of these concepts is as abstract as BELIEF (and more obscure), and there is no independent evidence that the child has constructed or uses any of them. There is good evidence that children think thoughts like, "SALLY BELIEVES THE MARBLE IS IN THE BASKET." But do children ever have thoughts like, "SALLY SEMANTICALLY EVALUATES A MENTAL REPRESENTATION EXPRESSING THE PROPOSITION THAT THE MARBLE IS IN THE BASKET"?

The difficulty of cashing in the claim that BELIEF is constructed encourages us to look elsewhere for an account. Rather than appealing to 'theory construction first, concept later,' we are exploring a 'concept first' approach—in which the reference of the concept is grounded by the operation of a mechanism and any 'theory construction' comes later.[3]

[3] For further discussion of these points see German and Leslie (2001) and Leslie (2000a) and for insightful discussion of the psychological nature of concepts see Fodor (1998).

Second, investigations of executive functioning in adults have stressed the fractionated nature of executive functioning (e.g., Goldman-Rakic, 1988, 1996; Shallice & Burgess, 1991), a point reflected in some recent developmental discussions (e.g., Carlson et al., 2002). Two types of fractionation are possible. First, EF may consist of a variety of different processes (Miyake et al., 2000), each of which provides a general resource for many domains. Second, EF may include processes that are dedicated to specific domains. Whereas type of EF fractionation is orthogonal to the expression/emergence issue, emergence accounts have so far assumed only domain general EF (e.g., Frye, Zelazo, & Palfai, 1995). However, 'theory of mind' EF may be wholly or partly distinct from general EF. Evidence from autism (Leslie & Roth, 1993; Leslie & Thaiss, 1992; Roth & Leslie, 1998), from early and later acquired brain damage (Fine, Lumsden, & Blair, 2001; Stone, Baron-Cohen, Calder, Keane, & Young, 2003; Varley, Siegal, & Want, 2001), and from neuro-imaging studies (e.g., Frith & Frith, 1999; Gallagher & Frith, 2003), suggest there is EF specific to the 'theory of mind' domain. Evidence from correlational studies (e.g., Carlson & Moses, 2001) do not establish the contrary because the observed correlations may reflect purely developmental features—e.g., maturational synchronies in distinct, but functionally similar, neural systems. Emergence accounts appear to assume something much stronger, namely, the exact same neural-processor for 'theory of mind' and non-'theory of mind' EF. As yet we see no evidence for this stronger assumption.

### 8.5. Epilogue

The fundamental principle of belief-desire reasoning is that people act to satisfy their desires in light of their beliefs. This principle is embodied in ToMM–SP's implicit mode of operation. Given a metarepresentational mechanism with the right MO, there is no need for the child to discover, reflect upon, or explicitly think about this principle. We have investigated and supported the following further aspects of this MO:

(a) provide candidate contents for belief attributions, minimally, a true-content, plus, if possible, plausible alternatives;
(b) assign initial salience/confidence levels to candidate contents, with highest level to 'sure' true-belief;
(c) review and adjust initial levels in light of specific circumstances;
(d) following review, select highest valued candidate.

Given this process, an account of early developmental change in reasoning about beliefs can be quite simple: namely, step (c) becomes more capable and accesses an increasing database.

We have suggested one way in which review effectiveness may increase, namely, better inhibitory control. Increased effectiveness will promote learning about false beliefs (Roth & Leslie, 1998), and through learning will come a larger database of circumstances under which inhibition should be initiated (for example, containers with unusual contents). At the root of this learning process is the ToMM, which frames the initial hypotheses for on-line heuristic belief-desire reasoning. We have argued elsewhere that ToMM may be a central module (Leslie, 1991, 1992, 1994a, 1994b; Leslie & Thaiss, 1992; Scholl & Leslie, 1999). If ToMM is a specialized and modular mechanism, SP is unlikely to be modular, though it may be dedicated to mentalizing. Both ToMM and SP may undergo developmental changes; but, given that it appears penetrable to instruction and knowledge, SP may be an especially important locus for developmental changes in heuristic mentalizing.

For kick-starting development, modular systems are ideal because they make minimal demands on general knowledge and general reasoning abilities (Leslie, 1986; Leslie & Keeble, 1987). Furthermore, by its very nature, the encapsulated knowledge of modules is implicit (inaccessible). Recent studies have shown implicit false-belief competence prior to 3 years of age (Clements & Perner, 1994; Onishi & Baillargeon, 2004; Garnham & Ruffman, 2001). Early implicit competence findings are expected on the ToMM account, while current theory-theories must make additional assumptions to remain consistent.

Finally, what about learning? Our approach assumes that ToMM has the potential to offer more than one candidate belief content to SP. By generating plausible hypotheses that, crucially, may involve false contents, the door to learning about false-belief is held open. ToMM may follow some simple rules of thumb. For example, in considering what Sally might think about X, ToMM may request information from central systems, such as, "What is true about X?" and "What did Sally last see/hear/access regarding X?" Such information may then be bound to the content slot of belief metarepresentations, forming hypotheses. At least for mundane, everyday situations, this should produce a short list of plausible initial hypotheses. These heuristics will be insufficient for the full range of effortful unhurried central reasoning about mental states that adults may undertake. But that is not the job of ToMM–SP. The main job of ToMM–SP is to allow the young brain to attend to invisible mental states and thereby start learning about them. But learning must start somewhere with something that is not itself learned. In conjunction with SP, ToMM functions as a mechanism and a conduit for learning about beliefs and other mental states.

**Acknowledgments**

We thank Jerry Fodor, Randy Gallistel, Rochel Gelman, Alvin Goldman, Steve Stich, Sarah-Jane Leslie, and especially Ori Friedman for helpful discussions and the latter, Susan Gelman, and two anonymous reviewers for detailed comments on earlier drafts. Preparation of this paper was supported by Grant BCS-0079917 from the National Science Foundation to A.M.L., and a UCSB Academic Senate Faculty Research Grant to T.P.G.

## Appendix A.

Basic protocols

***Avoidance Desire Task***

This is Sally. Look! She's got some food—it's a piece of fish. She wants to put the fish in a box. She is going to go inside to look for a box. [*goes inside*, *leaving fish behind*]

Here are two boxes. Let's look and see what is in them. In this box, there's a ball of wool. And in this other box, there's a ball of wool and there's also a poor, sick kitten. Sally does NOT want to give the poor little kitten the fish because it will make its tummy very sore. So she's going to go outside to get the piece of fish. She does NOT want to give the fish to the sick kitten. [*goes outside*] Why does she *not* want to? Yes, not to make the poor kitten worse!

| *True Belief* | *False-belief* |
|---|---|
| On her way back from getting the fish, look what Sally sees! The poor sick kitten crawls out of this box... and goes into this box. Did Sally see that? Yes! | Look what happens while she's gone! The poor sick kitten crawls out of this box... and goes into this box. Did Sally see that? No! |

Look, now Sally has the fish.
*Memory*: In the beginning, where was the kitten?
*Reality*: Where is the kitten now?

| | |
|---|---|
| *Know*: Does Sally know the kitten is in here? | *Think*: Where does Sally think the kitten is? |

*Prediction:* Which box will she go to with the fish?

***Mixed-Up Man task***

This is the "Mixed-Up Man". Do you know what he does? Every time he wants to do something, he does the *opposite*. If he wants an ice-cream, he eats a carrot! If he likes a cat, he pats a dog! If he wants something that is in here [box A], he looks in there [box B]. If he wants something in there [box B], he'll look in here [box A].

[*Man says*:] "Look, there's a piece of candy in this box. I love candy, so I'll look in *this* (opposite) box for the candy." [*take candy out of box*].

The Mixed-Up Man has a Mexican jumping bean. It jumps and wiggles around like this. Ok, one day, he puts his bean in this box. Then he goes on a walk. [*Exit*.]

| *True Belief* | *False-belief* |
|---|---|
| On his way back, look what he sees! The bean wiggles and jumps into the other box! [*moves*]. | While he's gone, look what happens! The bean wiggles and jumps into the other box! [*moves*]. |

**Appendix A** (*continued*)

| | |
|---|---|
| *Memory*: In the beginning, where was the bean? <br> *Reality*: Where is the bean now? | |
| *Know*: Does the man know his bean is in this box? | *Think*: Where does the man think his bean is? |
| *Prediction*: Where is he going to look for his bean? | |

## References


Astington, J. W., & Gopnik, A. (1991). Theoretical explanations of children's understanding of the mind. *British Journal of Develolpmental Psychology, 9*, 7–31.

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. Cambridge, MA: MIT Press.

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

Bartsch, K., & Wellman, H. M. (1995). *Children talk about the mind*. Oxford university 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., & Hix, H. R. (1998). The role of inhibitory processes in young children's difficulties with deception and false-belief. *Child Development, 69*, 672–691.

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.

Cassidy, K. W. (1998). Three- and 4-year-old children's ability to use desire-and belief- based reasoning. *Cognition, 66*, B1–B11.

Chomsky, N. A. (1957). *Syntactic structures*. The Hague: Mouton.

Chomsky, N. A. (1975). *Reflections on language*. New York, NY: Pantheon.

Clements, W. A., & Perner, J. (1994). Implicit understanding of belief. *Cognitive Development, 9*, 377–395.

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.

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.

Fine, C., Lumsden, J., & Blair, R. J. R. (2001). Dissociations between 'theory of mind' and executive functions in a patient with early left amygdala damage. *Brain, 124*, 287–298.

Fodor, J. A. (1992). A theory of the child's theory of mind. *Cognition, 44*, 283–296.

Fodor, J. A. (1998). *Concepts: Where cognitive science went wrong*. Oxford: Clarendon Press.

Frith, C. D., & Frith, U. (1999). Interacting minds—A biological basis. *Science, 286*, 1692–1695.

Frith, U., Morton, J., & Leslie, A. M. (1991). The cognitive basis of a biological disorder: Autism. *Trends in Neurosciences, 14*, 433–438.

Frye, D., Zelazo, P. D., & Palfai, T. (1995). Theory of mind and rule-based reasoning. *Cognitive Development, 10*, 483–527.

Gallagher, H. L., & Frith, C. D. (2003). Functional imaging of 'theory of mind'. *Trends in Cognitive Sciences, 7*, 77–83.

Garnham, W. A., & Ruffman, T. (2001). Doesn't see, doesn't know: is anticipatory looking really related to understanding of belief?. *Developmental Science, 4*, 94–100.
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.
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., & Nichols, S. (2003). Children's inferences about long and short causal chains. *Developmental Science, 6*, 514–523.
Goldman-Rakic, P. S. (1988). Development of cortical circuitry and cognitive function. *Child Development, 58*, 601–622.
Goldman-Rakic, P. S. (1996). The prefrontal landscape: Implications of functional architecture for understanding human mentation and the central executive. In A. Roberts & T. W. Robbins (Eds.), *The prefrontal cortex: Executive and cognitive functions* (pp. 87–102). London: Oxford Unversity Press.
Gopnik, A. (1993). How we know our minds: The illusion of first-person knowledge of intentionality. *Behavioral and Brain Sciences, 16*, 1–14.
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.
Hale, C. M., & Tager-Flusberg, H. (2003). The influence of language on theory of mind a training study. *Developmental Science, 6*, 346–359.
Leslie, A. M. (1986). Getting development off the ground: Modularity and the infant's perception of causality. In P. van Geert (Ed.), *Theory building in development* (pp. 405–437). Dordrecht: North Holland.
Leslie, A. M. (1987). Pretense and representation: The origins of theory of mind. *Psychological Review, 94*, 412–426.
Leslie, A. M. (1991). The theory of mind impairment in autism: Evidence for a modular mechanism of development?. In A. Whiten (Ed.), *Natural theories of mind: Evolution, development and simulation of everyday mindreading* (pp. 63–78). Oxford: Blackwell.
Leslie, A. M. (1992). Pretense, Autism, and the Theory of Mind module. *Current Directions in Psychological Science*, 1, 18–21. (Reprinted in R.P. Honeck (Ed.), *Introductory Readings for Cognitive Psychology*, 3rd edition, Guilford, CT: Dushkin/McGraw-Hill, 1997).
Leslie, A. M. (1994a). *Pretending and believing*: Issues in the theory of **ToMM**. *Cognition*, 50, 211–238. (Reprinted in J. Mehler & S. Franck (Eds.), *COGNITION on cognition*, pp. 193–220, Cambridge, MA: MIT Press, 1995).
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.
Leslie, A. M. (2000a). How to acquire a representational theory of mind. In D. Sperber (Ed.), *Metarepresentations: A multidisciplinary perspective* (pp. 197–223). Oxford: Oxford University Press.
Leslie, A. M. (2000b). 'Theory of mind' as a mechanism of selective attention. In M. Gazzaniga (Ed.), *The new cognitive neurosciences*, (2nd ed., 1235–1247). Cambridge, MA: MIT Press.
Leslie, A. M., & Keeble, S. (1987). Do six-month-old infants perceive causality?. *Cognition, 25*, 265–288.
Leslie, A. M., & Polizzi, P. (1998). Inhibitory processing in the false-belief task: Two conjectures. *Developmental Science, 1*, 247–254.
Leslie, A. M., & Roth, D. (1993). What autism teaches us about metarepresentation. In S. Baron-Cohen, H. Tager-Flusberg, & D. Cohen (Eds.), *Understanding other minds: Perspectives from autism* (pp. 83–111). Oxford: Oxford University Press.
Leslie, A. M., & Thaiss, L. (1992). Domain specificity in conceptual development: Neuropsychological evidence from autism. *Cognition, 43*, 225–251.
Mitchell, P. (1994). Realism and early conception of mind: A synthesis of phylogenetic and ontogenetic issues. In C. Lewis & P. Mitchell (Eds.), *Children's early understanding of mind* (pp. 141–172). New York: Cambridge University Press.
Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*, 49–100.

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

Onishi, K., & Baillargeon, R. (2004). False-belief understanding in infancy. Paper presented at symposium on Non-verbal measures of theory of mind, *Biennial International Conference on Infant Studies*, Chicago, IL, May 2004.

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. (1999). Development of theory of mind and executive control. *Trends in Cognitive Science, 3*, 337–344.

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

Perner, J., & Wimmer, H. (1985). "John thinks that Mary thinks that...": Attribution of second-order false beliefs by 5-to 10-year-old children. *Journal of Exprimental Child Psychology, 39*, 437–471.

Posner, M. I., & Cohen, Y. (1984). Components of visual orienting. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance, X* (pp. 531–556). Hillsdale, NJ: Erlbaum.

Rafal, R., & Henik, A. (1994). The neurology of inhibition: Integrating controlled and automatic processes. In D. Dagenbach & T. H. Carr (Eds.), *Inhibitory processes in attention, memory and language* (pp. 1–51). New York: Academic Press.

Richardson, J. T. E. (1990). Variants of chi-square for $2 \times 2$ contingency tables. *British Journal of Mathematical and Statistical Psychology, 43*, 309–326.

Riggs, K. J., & Mitchell, P. (2000). Making judgements about mental states: Processes and inferences. In P. Mitchell & K. J. Riggs (Eds.), *Children's reasoning and the mind* (pp. 1–10). Hove, East Sussex: Psychology Press.

Roth, D., & Leslie, A. M. (1998). Solving belief problems: Toward a task analysis. *Cognition, 66*, 1–31.

Russell, J., Mauthner, N., Sharpe, S., & Tidswell, T. (1991). The 'windows task' as a measure of strategic deception in preschoolers and autistic subjects. *British Journal of Developmental Psychology, 9*, 331–349.

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.

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.

Shallice, T., & Burgess, P. (1991). Higher-order cognitive impairments and frontal lobe lesions in man. In H. S. Levin, H. M. Eisenberg, & A. L. Benton (Eds.), *Frontal lobe function and dysfucntion* (pp. 125–138). Oxford: Oxford University Press.

Siegal, M., & Beattie, K. (1991). Where to look first for children's knowledge of false beliefs. *Cognition, 38*, 1–12.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review, 63*, 129–138.

Stone, V. E., Baron-Cohen, S., Calder, A., Keane, J., & Young, A. (2003). Acquired theory of mind impairments in individuals with bilateral amygdala impairments. *Neuropsychologia, 41*, 209–220.

Surian, L., & Leslie, A. M. (1999). Competence and performance in false-belief understanding: A comparison of autistic and 3-year-old children. *British Journal of Developmental Psychology, 17*, 141–155.

Varley, R., Siegal, M., & Want, S. C. (2001). Severe impairment in grammar does not preclude theory of mind. *Neurocase, 7*, 489–493.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory mind development: The truth about false-belief. *Child Development, 72*, 655–684.

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