THE NEW COGNITIVE NEUROSCIENCES

Second Edition

Michael S. Gazzaniga, *Editor-in-Chief*

0-262-07195-9

A Bradford Book

The MIT Press
Cambridge, Massachusetts
London, England
2000

85 "Theory of Mind" as a Mechanism of Selective Attention

ALAN M. LESLIE

ABSTRACT A key component of human intelligence is our ability to think about each other's mental states. This ability provides an interesting challenge for cognitive neuroscience attempts to understand the nature of abstract concepts and how the brain acquires them. Research over the past 15 years has shown that very young children and children of extremely limited intellectual ability can acquire mental state concepts with ease. Children with Kanner's syndrome have severe difficulty using these concepts, despite relatively great experience and ability. These discoveries have led to the development of the first information processing models of belief-desire reasoning.

The term "theory of mind" was coined by David Premack (Premack and Woodruff, 1978) to refer to our ability to explain, predict, and interpret behavior in terms of mental states, like *wanting, believing,* and *pretending.* Because the behavior of complex organisms is a result of their cognitive properties—their perceptions, goals, internal information structures, and so on—it may have been adaptive for our species to develop some sensitivity to these properties. The capacity to attend to mental state properties is probably based on a specialized representational system and is evident even in young children.

The term "theory of mind" is potentially misleading. It might suggest that the child really has a *theory* or that the child has a theory of *mind* as such. Although there are some writers who hold such views (Perner, 1991; Gopnik and Meltzoff, 1997; Gopnik and Wellman, 1995), I assume simply that the child is endowed with a representational system that captures cognitive properties underlying behavior. To better see what is meant by "theory of mind" ability, consider the following scenario (figure 85.1). Sally has a marble that she places in a basket and covers, and then departs. While she is gone, Ann removes the marble from the basket and places it in the box. A child to whom this scenario is presented then is asked to predict where Sally will look for her marble when she returns. To correctly predict Sally's behavior, it is necessary to take into account both Sally's desire for the marble and Sally's belief concerning the location of the marble. In this scenario, Sally's belief is rendered false by Ann's tampering. Therefore, to succeed on this task, the child must attribute to Sally a belief that, from the attributer's point of view, is false.

There have been two major discoveries concerning the false-belief problem in figure 85.1. First, Wimmer and Perner (1983), using a somewhat more complex version of the task, found that the majority of 6-year-olds already could pass it, whereas Baron-Cohen and associates (1985), using the version depicted, found that the majority of 4-year-olds could succeed. Subsequently, a large number of studies have confirmed this finding: Whether predicting behavior or reporting where Sally thinks the object is, normally developing children typically solve the problem shortly after the fourth birthday. The second major finding is that autistic children typically fail to solve this task despite mental ages (MAs) well in excess of 4 years, whereas other disabled children—for example, those with Down syndrome—can succeed (Baron-Cohen, Leslie, and Frith, 1985). These two findings raise the following deeply challenging problem for the theorist of cognitive development. How is the young brain able to attend to mental states when mental states cannot be

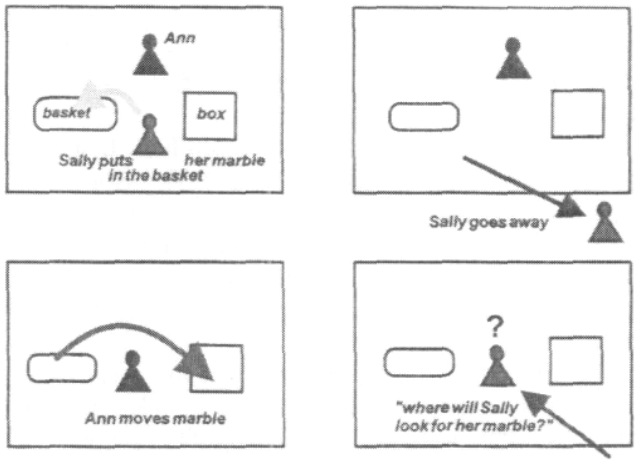

FIGURE 85.1 Illustration of the standard "Sally and Ann" false-belief task given to children to test their ability to attribute beliefs to other people and to calculate the contents of those beliefs correctly. (After Baron-Cohen, Leslie, and Frith, 1985.)

ALAN M. LESLIE Department of Psychology and Center for Cognitive Science, Rutgers University, Piscataway, NJ.

FIGURE 85.2 The 2-year-old brain can attend to behavior and infer the mental state from which the behavior issues. To do this requires the development of the "M-representation." See this page for further explanation. (After Leslie, 1994.)

seen, heard, or felt? I call this the *fundamental* problem of "theory of mind" because if the child cannot attend to mental states, then how can he or she learn about them?

Previously (Leslie, 1987), I proposed a general answer to the fundamental question of how the young brain can attend to mental states: Attend to behavior and infer the mental state from which the behavior arises. For example, the 2-year-old child watches as mother talks to a banana. If the child were limited to representing simply the mother's behavior then he or she would be unable to recover the significance of the mother's behavior. This he or she can do only by inferring that the mother is pretending that the banana is a telephone (figure 85.2). In fact, 2-year-olds are quite capable of this feat (Harris and Kavanaugh, 1993; Leslie, 1987, 1994).

For the young brain to move attention from behavior to the mental state from which the behavior issues, appropriate processing mechanisms must deploy a system of representation capable of representing mental states. I call this system of representation the *M-representation,* and the associated processing mechanisms the *theory of mind mechanism* (ToMM).

The M-representation provides agent-centered descriptions of behavior using three-place relations that make explicit four kinds of information. The first piece of information specifies the agent involved. The second piece of information identifies an "informational relation" or attitude that the agent holds. The third piece of information identifies an aspect of the world that anchors the agent's attitude, and the final piece of information identifies the content of the agent's attitude. For example, in the agent-centered description of the mother's behavior shown in figure 85.2, the M-representation shows the agent, **mother,** holding the attitude, **pretends-true,** toward the content, **"it is a telephone,"** with regard to the **banana.** The M-representation system is highly flexible. The present example can be extended easily to cover: own pretend play, by having the child process an M-representation which identifies the agent as self (I **pretend-true** of **the banana "it is a telephone");** different mental states, including false beliefs **(mother believes-true of the banana "it is a telephone");** different anchors; and different contents. Forming and processing an M-representation requires the brain to integrate information from a number of very different sources.

These ideas led my colleagues and me to develop a neuropsychological perspective on autism, a perspective that also has helped us to understand the normal development of social intelligence (Baron-Cohen, 1995; Frith, Morton, and Leslie, 1991; Happé, 1995; Leslie, 1987, 1991, 1992, 1994; Leslie and German, 1995; Leslie and Roth, 1993; Leslie and Thaiss, 1992; Roth and Leslie, 1998; see also chapter 87). According to the model, the M-representation is deployed by a dedicated processor, ToMM. The ToMM is a specialized component of social intelligence providing the time-pressured, on-line intentional interpretations of behavior that are necessary for an agent to take part effectively in conversations and other real-time social interactions. The ToMM is a mechanism of selective attention, it operates postperceptually, it operates spontaneously whenever an agent's behavior is attended, it is domain specific, and it is subject to dissociable damage. In the limit, the ToMM may be modular. The ToMM employs a proprietary representational system, namely the M-representation. The ToMM is hypothesized to form the specific basis of our capacity to acquire "theory of mind." Finally, the ToMM is damaged in autism (Kanner's syndrome), resulting in the core signs of that neurodevelopmental disorder.

Background assumptions about autism

Autism is a disorder affecting at least 4 or 5 in 10,000 births; approximately 75% of those affected are mentally retarded. The evidence is overwhelming that the disorder has a biological etiology (Gillberg and Coleman, 1992) and is most likely genetic in origin (see chapter 87; also see Bailey et al., 1995).

At present, autism is diagnosed on behavioral grounds, including impaired social skills, language delay, lack of pretend play, and stereotypes, with onset before

36 months of age (American Psychiatric Association, 1994). Large-scale epidemiological studies by Wing and Gould (1979) showed that autistic children suffer a "triad of impairments" relative to nonautistic mentally retarded children matched on mental age. The triad of impairments includes social incompetence, poor verbal and nonverbal communicative skills, and a lack of pretend play. Although approximately 25% of children with autism are not mentally retarded, they still show the "triad of impairments" compared with their peers. This suggests that the triad, although central to the syndrome of autism, is not the result of general mental retardation but reflects a more specific impairment at the cognitive level (Leslie, 1987). Because "theory of mind" abilities underlie human social competence, communication, and pretending, the autistic triad might be the result of an impaired ToMM. These speculations led to the prediction that autistic children would be specifically impaired in their understanding of beliefs in other people.

Investigating the theory of mind mechanism hypothesis: Initial phase

To test the predicted impairment in belief understanding, Baron-Cohen, Leslie, and Frith (1985) studied three groups of children: normally developing 4-year-olds, children with autism, and children with Down syndrome. Subjects were tested on the Sally and Ann false-belief task (figure 85.1). To allow a conservative test of the hypothesis, the autistic children were older (12 years) and thus more experienced than the other two groups (10 years and 4 years) and had a higher mean IQ (82) than the children with Down syndrome (64). After an experimenter explained the scenario with the aid of props, subjects were asked three questions: a Memory question, "In the beginning, where did Sally put her marble?"; a Reality question, "Where is the marble now?"; and a Prediction test question, "Where will Sally look for her marble?"

The results were striking. Eighty-five percent of the normally developing children and 86% of the children with Down syndrome attributed Sally a false belief and predicted that she would look in the basket. Only 20% of the autistic children predicted Sally's behavior in this way, failing as a group to show their advantage in age, experience, and ability.

Leslie and Frith (1988) replicated these findings with a group of autistic children with mean verbal MAs of 7 years, 2 months, comparing them to MA-matched specific language impaired (SLI) children. All the SLI children passed the task; by contrast, only 28% of the autistic children passed. Leslie and Frith also showed that although autistic children were perfect in a "line of sight" task, they performed poorly in a test of "seeing leads to knowing." Perner and associates (1989) investigated a second false-belief task with autistic children. In this task, the child is shown a container for a well-known candy and asked, "What's in here?" After the child names the candy, the container is opened and the child is shown that it contains only a pencil. The pencil then is replaced and the container again closed. The child is told that when his or her friend comes in, the friend too will be shown the container and asked what is inside. The child then is asked what the friend will say. The results on this task for normally developing children closely follow those obtained from the Sally and Ann task: typically, 3-year-olds fail to predict behavior by attributing a false belief, whereas 4-year-olds typically succeed. Perner and colleagues (1989) found that almost 100% of MA-matched SLI children passed the candies task, whereas 83% of their able autistic group failed.

These initial findings have been replicated and extended by different laboratories around the world (Baron-Cohen, 1995; Mitchell, Saltmarsh, and Russell, 1997; Naito, Komatsu, and Fuke, 1994; Ozonoff, Pennington, and Rogers, 1991; Prior, Dahlstrom, and Squires, 1990; Reed, 1994; Sodian and Frith, 1992; Tager-Flusberg, 1992). Autistic children are impaired in their understanding of beliefs relative to normal developmental milestones, relative to their own level of general intellectual functioning, and relative to other syndromes of mental retardation and language impairment. This pattern is consistent with impairment to their ToMM.

A key question and a key finding

How do we know that the failure of autistic children on false-belief tasks is not due to an impairment in general processing or general reasoning? It is easy to think of nonspecific impairments that would impact on these tasks, for example, impaired working memory, poor executive function, limited abstract reasoning, difficulties with counterfactual reasoning, or other nameless processing factors impaired in critical combinations.

To answer this question, a task that closely parallels the general problem-solving structure of false-belief tasks,but without engaging mental state concepts, would be useful. Zaitchik (1990) devised just such a task, the so-called "photographs" task, in which Sally is downsized, replaced by "hi-tech," namely a Polaroid camera. Sally's belief is replaced by a photograph: a mental representation is replaced by a public representation.

Because photographs and other pictures are easily attended to, can be picked up, pointed to, discussed with

mother, and almost always are out of date, they should have a marked advantage in the development of the child over invisible, intangible, immaterial beliefs.

The task begins by ensuring that preschoolers understand the basics of the operation of the camera. After training, the children are asked to take a photograph of a toy cat sitting on a chair (figure 85.3). When the photograph emerges from the camera, the experimenter places it face down on a table. The child does not get to see the photograph; after all, the child did not get to see Sally's belief! The cat then is moved from the chair and placed on the bed. The child is asked the usual control questions, "When you took the photograph, where was the cat? Where is the cat now?" Finally, the child is asked the crucial test question, "In the photograph, where is the cat?"

When Zaitchik gave this task to preschoolers, the results resembled those obtained from the false-belief task: 3-year-olds typically failed, answering the test question with the current location of the cat, whereas 4-year-olds typically passed.

Leslie and Thaiss (1992) adapted this task for use with autistic children and compared their performance with that of normally developing 4-year-olds. Two standard false-belief tasks, the Sally and Ann and candies tasks, were given, along with two photographs tasks—the aforementioned task and a second task in which the photographed object subsequently is replaced with a different object. In this latter task, the test question is, "In the photograph, what is on the chair?" This asks for the identity of an object, as does the test question in the candies task.

The results showed that most of the normally developing 4-year-olds passed both the out-of-date belief tasks and their equivalent out-of-date photograph tasks. Although their performance on the belief and photograph tasks did not differ significantly, for children who passed only one of the tasks, there was a tendency to pass false belief. This tendency also was found by Zaitchik (1990) in her three experiments that allowed the comparison. Together with the pair of tasks from Leslie and Thaiss (1992),these five studies show the same small advantage for out-of-date beliefs over photographs, *using closely parallel task structures.* Experiment-wise, the effect is reliable *(p = .032)*, with normally developing children finding false beliefs slightly easier than out-of-date photographs.[1]

If only general learning mechanisms are involved, it is surprising that photographs and other pictures are not easier to learn about than beliefs. Moreover, despite attempts to put a brave face on it (Perner, 1995; Leekam and Perner, 1991), these findings are a particular embarrassment for accounts in which the child comes to understand belief by discovering a "theory" about mental states, namely, the theory that "*mental states are representations"* (Perner, 1991). The failure to find a large advantage for pictures over beliefs supports the idea that something in the young brain compensates for the invisibility of belief. Apparently, if anything, beliefs are easier, not harder, to learn about. This highlights the proposed role of the ToMM as a mechanism that directs attention to otherwise unattendable mental states and thus promotes learning.

The results from the autistic subjects in the study by Leslie and Thaiss (1992) were strikingly different. Autistic children showed their characteristic poor performance on both false-belief tasks coupled with near perfect performance on the equivalent photographs tasks, reversing the pattern found in normally developing children.

Leslie and Thaiss ran a further study in which the camera and photograph were replaced by a "map." Subjects were familiarized with a simple diagrammatic map of a doll's house and were trained in how a puppet, placed on a piece of furniture, could have its position marked on the map with a colored sticker. During testing, the experimenter placed a puppet on the bed and placed a sticker on the map to show where the dog was. Again, the child did not get to see the marked map. The doll then was moved from the bed onto the toy box. After the usual control questions, the child was asked "In the map, where is the doll?" Again, the autistic children fared better on the public representation task than on false belief, whereas the normally developing children showed the opposite pattern.

Figure 85.4 puts these results together. It shows how the autistic children performed with the normal profile on public representation tasks, only with greater success—*as they should,* given their age and ability advantage over the other group. When a belief task enters the comparison, a crossover emerges.

These results help rule out a whole class of explanation for the poor performance of autistic children on false-belief tasks. For example, if autistic children perform poorly on false-belief tasks because of limited working memory, because of poor executive function, or because of impaired counterfactual reasoning, then why did the photographs/maps tasks not also demand these things? Similarly, for impaired event memory, for poor attention shifting, for poor mental imagery, and for other "general" impairments, it is hard to see why false-belief problems require the favored resource while other representation tasks do not. These findings challenge accounts that rely on an impairment in a general capacity.

The double dissociations in figure 85.4 suggest that although autistic children possess the general problem-solving resources required by the false-belief task, they are impaired in a specific representational competence,

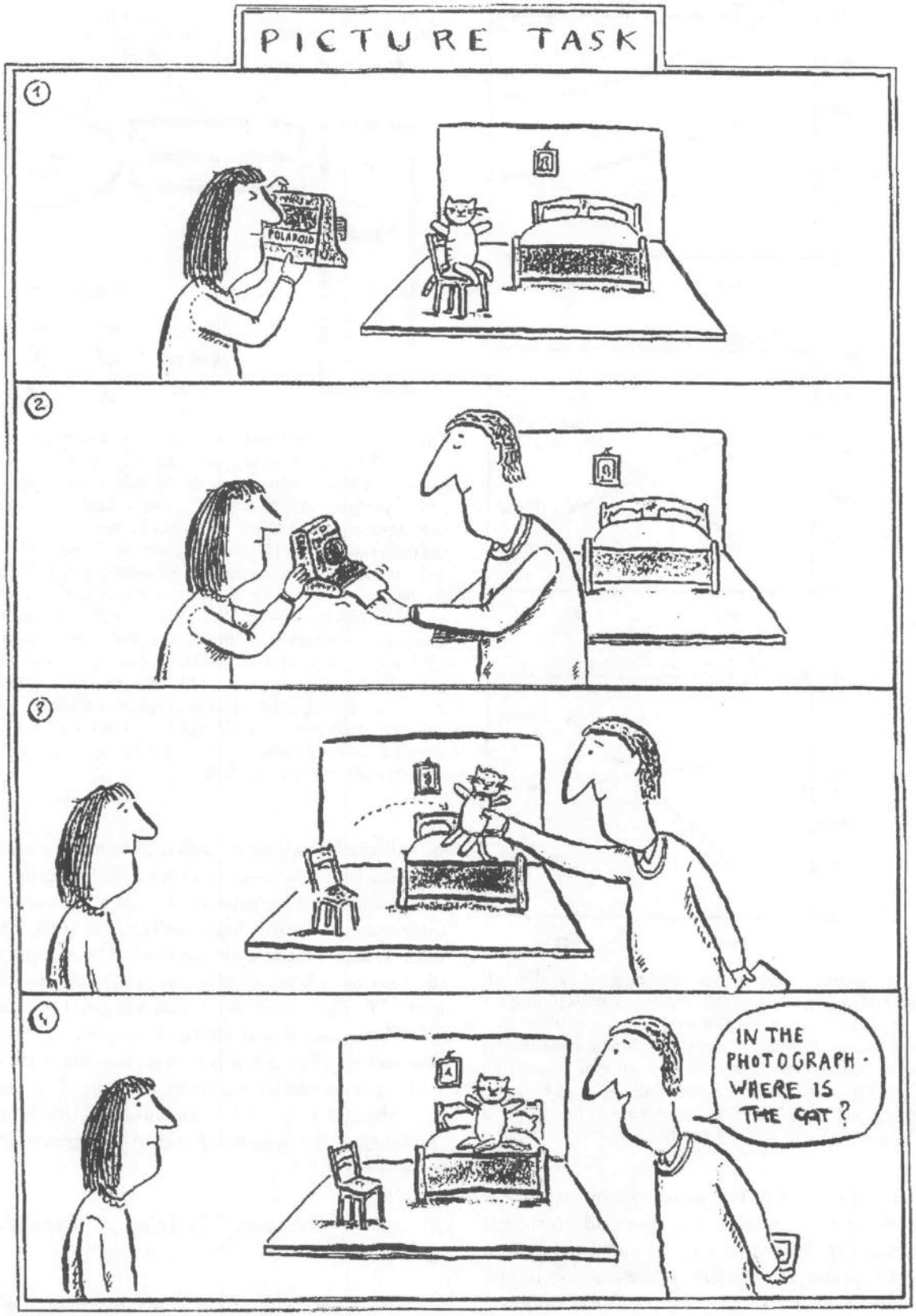

FIGURE 85.3 The out-of-date photograph task. (Reproduced from Happé, 1995, by permission of the artist, Axel Scheffler.)

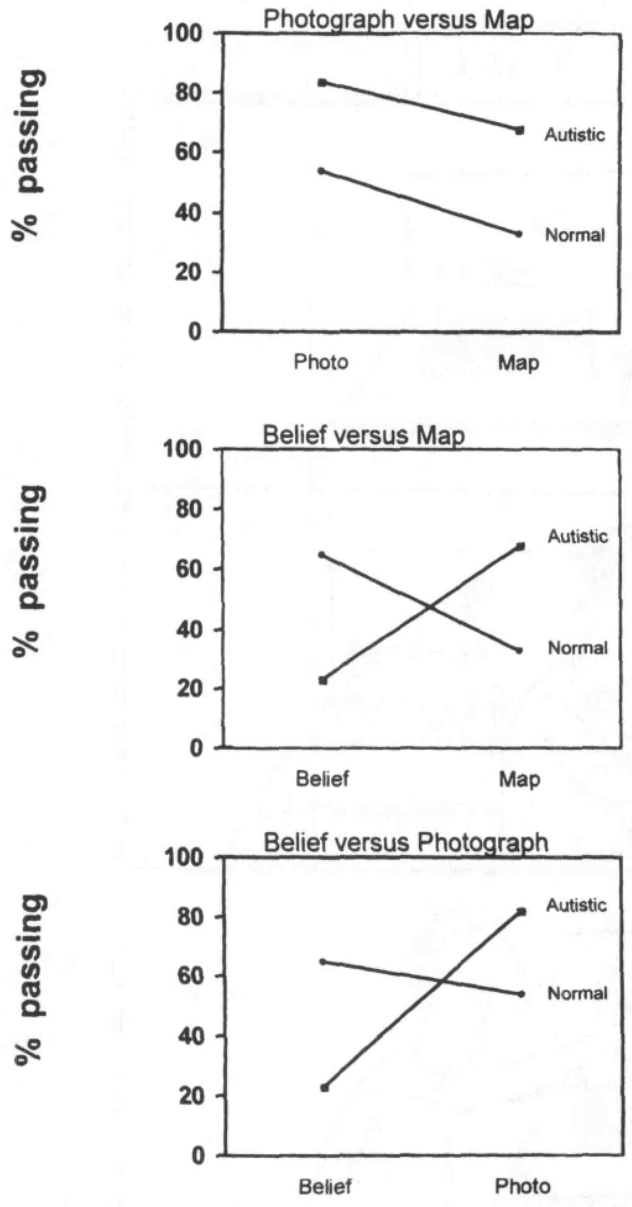

FIGURE 85.4 Autistic children show their age and ability advantage over normally developing 4-year-old children on tasks testing understanding of out-of-date public representations (top panel). When tasks with the same general structure but testing understanding of the mental state *belief* are introduced, autistic performance collapses, revealing a double dissociation between understanding public and mental representations (bottom two panels). (Data from Leslie and Thaiss, 1992.)

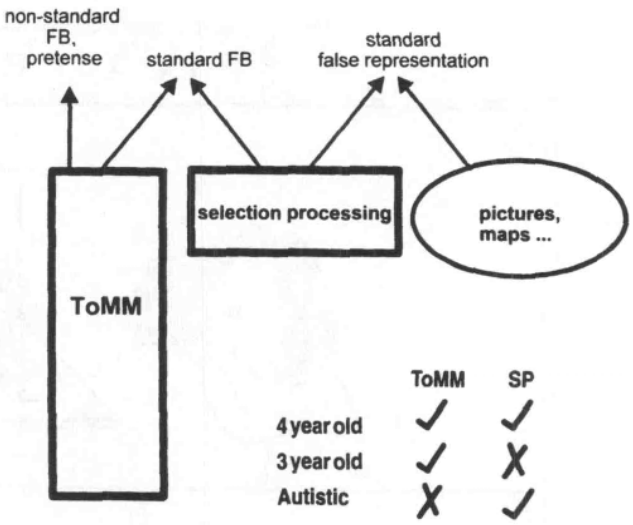

FIGURE 85.5 The theory of mind mechanism-selection processing (ToMM-SP) model of "theory of mind" development. Some problems require only the ToMM, for example, recognizing pretense and "modified" nonstandard false-belief tasks that typically are passed by normally developing 3-year-old children but are hard for autistic children. Standard false-belief tasks require the cooperation of SP with the ToMM, typically are failed by 3-year-olds and by autistic children, but typically are passed by 4-year-olds. Public representation tasks with the same task structure as standard false-belief tasks also require SP but do not involve the ToMM; these tasks are passed by normally developing 4-year-olds and autistic children. Normally developing children have an intact ToMM, but younger children have only weak SP. Autistic children are a mirror image of 3-year-olds with adequate SP but an impaired ToMM. (After Leslie and Thaiss, 1992.)

for example, the ToMM. The picture for normally developing children is crucially different and consistent with the idea that those who fail false-belief tasks do so because of limitations in general resources. Leslie and Thaiss proposed that the ToMM alone is not sufficient for the standard false-belief task, which requires a further mechanism. They called this extra mechanism selection processing (SP). They argued that the ToMM by default attributes a belief with a content that reflects current reality. To succeed in a false-belief task, this default attribution must be *inhibited* and an alternative nonfactual content for the belief selected instead. Standard photograph tasks also demand SP. The normal 3-year-old fails both kinds of task because he or she has only weak SP. The successful 4-year-old has both an intact ToMM and sufficiently strong SP to pass standard tasks. The autistic child is a mirror image of the 3-year-old in so far as he or she has sufficiently strong SP (to pass photographs and maps) but an impaired ToMM. Figure 85.5 summarizes the ToMM-SP model of "theory of mind" development.

Examining failure on false belief: A second phase of research

Thus far, autistic children have been compared with normally developing 4-year-olds and nonautistic mentally retarded children who *pass* standard false-belief tasks. This work has established a specific impairment in "the-

ory of mind" abilities in autistic children. More recently, we have been studying the reason for autistic failure by comparing autistic children with normally developing 3-year-olds who also *fail* standard false-belief tasks. Do these two groups fail for the same reasons?

A number of studies are showing that these groups fail for different reasons. In the first such study, Roth and Leslie (1991) modified a false-belief task so that it became easier for 3-year-olds. Although most of the 3-year-olds in this task successfully attributed a false belief, most of the older autistic subjects did not. Roth and Leslie (1998) found differences between normally developing 3-year-olds and older autistic subjects on a "seeing leads to knowing" task. In a further experiment, Roth and Leslie (1998) showed that a modified false-belief task was easier than a standard false-belief task for 3-year-olds but not for autistic subjects, and that selection processing demands were a limiting factor on 3-year-old performance but not on autistic performance.

A number of modifications to the standard false-belief task are known to help 3-year-olds achieve better performance. The most minimal modification to the standard false-belief task that helps 3-year-olds is to ask, "Where will Sally look *first* for her marble?" Siegal and Beattie (1991) found that the addition of the single word "first" dramatically improved performance in a task in which children are told explicitly what Sally thinks. Surian and Leslie (1999) applied this minimal modification to a standard Sally and Ann task. One potential problem with asking where Sally will look first is that children simply might respond with where the object had been placed *first.* They then would appear to pass the task without ever considering Sally's belief. Or children might assume, on being asked about a first look, that there will be a series of looks culminating in success and that the first look will therefore be a *failing* look. Again, responding with a failing look, children will appear to pass the task without ever considering Sally's belief. To control for these possibilities, Surian and Leslie tested children on a control task in which Sally does not go away but instead watches while Ann moves the marble from the basket to the box. In this case, Sally knows where the marble is. If children in this condition follow either of the placed-first or failing-look strategies, they again will indicate the empty location; however, in the true-belief condition, such a response is wrong.

Normally developing children approximately 3 years and 9 months of age were tested in one of four conditions: a standard false-belief condition; an equivalent "standard" true-belief condition in which Sally stays and watches; a "look first" false-belief condition; and an equivalent "look first" true-belief condition. Only 30% of children in the standard false-belief condition passed,

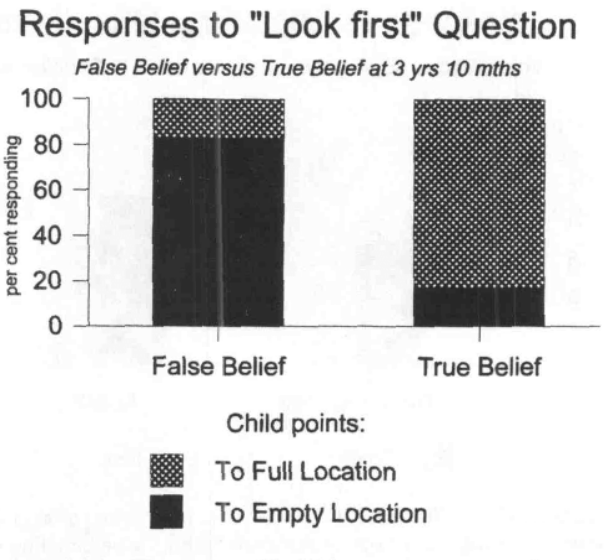

FIGURE 85.6 Three-year-old children respond correctly to a "look first" prediction question in both false-belief and true-belief tasks, although correct responses are opposite in the two tasks. The "look first" modification therefore helps 3-year-old children to calculate belief. (After Surian and Leslie, 1999.)

a typical result, whereas all the children passed the corresponding true-belief version. In the "look first" conditions, 83% of the children passed false belief, whereas the same proportion passed the true-belief equivalent. The children in the latter two groups produced opposite responses when asked "Where will Sally look first for her marble?" depending on the belief status of Sally (figure 85.6).

The "look first" question helps younger children calculate a false belief. How does it do this? One possibility suggested by Siegal and Beattie (1991) is that it helps younger children recognize the questioner's intention to ask about a belief rather than about reality. This may well be correct, but it does not indicate how it helps younger children do this, nor does it indicate why slightly older children do not need such help. Surian and Leslie (1999) suggest that by directing children's attention to the first location of the object or to the possibility of failing looks, the question increases the salience of the first location *as the possible content* of Sally's belief. This increased salience of the nonfactual content relative to the default reduces the need for inhibition, and thus, the task places less load on SP

Surian and Leslie (1999) carried out a second experiment to determine whether the "look first" question also would help autistic children with false belief. In this study, a "think" question was asked before the usual control questions. After the control questions, the children were asked the "look first" version of the "prediction" question. With this design, the same child can be

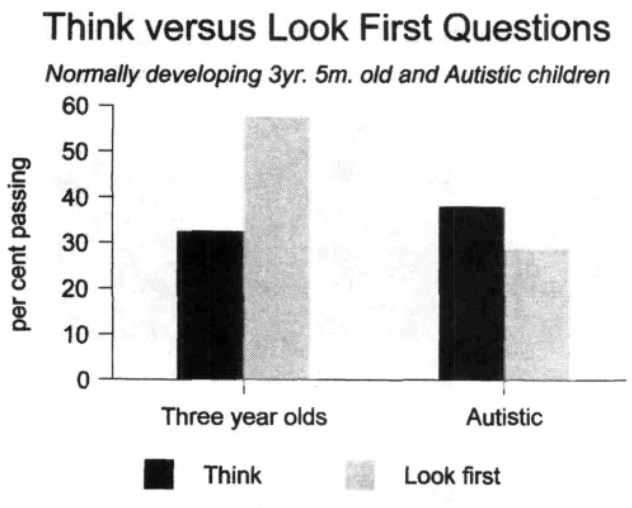

FIGURE 85.7 The "look first" question improves younger 3-year-old performance over a standard "think" false-belief question. No such improvement is seen in older autistic children. (After Surian and Leslie, 1999).

asked a standard task "think" question followed by a nonstandard "look first" question. A comparison group of normally developing children of approximately 3 years and 5 months of age also was tested in this way. The normally developing children again were helped by the "look first" question, although not as much as before, presumably because they were younger. The children with autism, by contrast, were not helped at all (figure 85.7).

In summary, normally developing 3-year-olds and older autistic children fail false-belief tasks for different reasons. There is growing evidence that normally developing children's performance on false-belief problems is limited by processing resources rather than by an inability to represent belief states in others. These processing resources increase gradually over the preschool period and by the time the child is a little older than 4 years of age usually are sufficient to allow success on standard tasks. Older children with autism fail false-belief tasks for different reasons, apparently reflecting an impaired capacity to acquire normal "theory of mind" knowledge and skills. The emerging pattern supports the ToMM-SP model.

Inhibition in belief-desire reasoning

So far we have been considering why children fail false-belief tasks. It is equally important to develop models of how children pass these tasks. One requirement is the ability to represent the right kinds of information: information about agents, attitudes, anchors, and contents tied together in the relational structure modeled by the M-representation. But simply being able to deploy concepts, like *pretend*, *desire*, and *believe*, does not guarantee that the child is able to solve particular "theory of mind" problems or already knows particular "theory of mind" facts.

We saw earlier that default belief attributions need to be inhibited to solve a false-belief problem and suggested that producing and controlling this inhibition may be a problem for young children. Performance changes due to the maturation of prefrontal cortex may be ubiquitous in development (Diamond, 1988; Goldman-Rakic, 1987), and Carlson and associates (1998) provide independent evidence of inhibitory involvement in the development of "theory of mind" skills.

It is useful to understand why belief attribution has a default bias. If *desires* set an agent's goals, *beliefs* inform the agent about the state of the world. A belief that misinforms an agent is a useless, even dangerous, thing: beliefs *ought* to be true. Therefore, the best guess strategy for the naive belief attributer is to assume that an agent's beliefs *are* true. Apparently, this is the strategy followed by the 3-year-old. However, false-belief tasks require that the default strategy be over-ridden. According to the ToMM-SP model, to do so requires inhibition of the prepotent attribution. The older child's success shows that he manages this inhibition.

In a standard false-belief task, there are essentially two possible locations, the basket and the box, to which Sally's belief about the marble might refer or which might be targets of Sally's desire. The default belief attribution draws attention to one of these locations, namely, the current location of the object. To successfully solve the false-belief problem, the brain must disengage attention from this target and shift to the false-belief target. Inhibitory brain processes appear to be involved in other kinds of attention shifting, for example, in shifting covert visual attention (Posner and Presti, 1987; Rafal and Henik, 1994). According to the ToMM-SP model, belief-task target shifting also requires inhibitions.

Leslie and Polizzi (1998) tested the belief inhibition hypothesis by following up a finding of Cassidy (1995). Cassidy found that when the desire in a standard false-belief task is negative rather than positive, then 4-year-old children perform poorly. Typically, in false-belief task scenarios, Sally wants the target object; but in Cassidy's task, Sally did *not* want to find the target. Leslie and Polizzi argued that the critical feature was not negation as such but whether the negation produced target shifting. Suppose a protagonist has a desire for whichever location does not have property X, and the only way to identify the NOT(X) location is to first identify the X location, and then choose the other one, that is, NOT(X). Identifying the protagonist's desire target

this way involves the brain in target shifting. For example, Sally has a box and a basket that both contain some wool. She does not want to put a fish in the basket because there is a sick kitten nestling in the wool there (the kitten might get worse if it eats the fish). To identify where Sally wants to put the fish, one first identifies which location has the kitten, then, because this is *not* what Sally wants, one shifts from this location to the alternative. This creates a target-shifting desire.

Leslie and Polizzi (1998) pointed out an interesting feature of inhibition models of belief-desire reasoning. With true belief and positive desire, there is no target shifting. A false belief (with positive desire) involves a single target-shifting inhibition; so does a target-shifting desire (with true belief). However, when a false belief is combined with a target-shifting desire to produce a double inhibition task, the two inhibitions cannot simply sum their inhibitions because this will produce the wrong answer. Working through figure 85.8 will clarify this last point.

The four panels in figure 85.8 correspond to four kinds of behavior prediction task. The first is the simple true-belief plus positive-desire task: Sally wants the object and knows where it is. The pointing hand represents a mental index that the brain uses to indicate the target of Sally's belief and desire, and therefore the answer to the prediction task. In this model, belief and desire targets are identified in parallel. The next panel is the standard false-belief task. Here the first belief-desire index again indicates the location that contains the object because the initial belief attribution always is the default true belief, that is, where someone *should* think the object is. But in this task, the protagonist's belief is false and so, to succeed, the subject must inhibit this index. In the second panel, the inhibition is visualized by the "inhibition arm" reaching in to weaken the index. Because the initial target is inhibited, the index moves across to the alternate location, yielding the correct prediction that the protagonist will look in the empty location.

The third panel in figure 85.8 shows a true-belief with target-shifting desire. Again, the initial belief index shows the default true-belief location. The initial desire index shows the same target because the protagonist desires the NOT(X) location and the only way to identify that is to first identify the X location, then cancel it in favor of NOT(X). So in this panel, the inhibition arm inhibits the desire target and again the index moves to the empty location, generating the correct prediction.

The final panel shows how to predict behavior when the protagonist's belief is false and desire "negative." Once again, the belief-desire index initially is placed against the full location. But now two inhibitions must

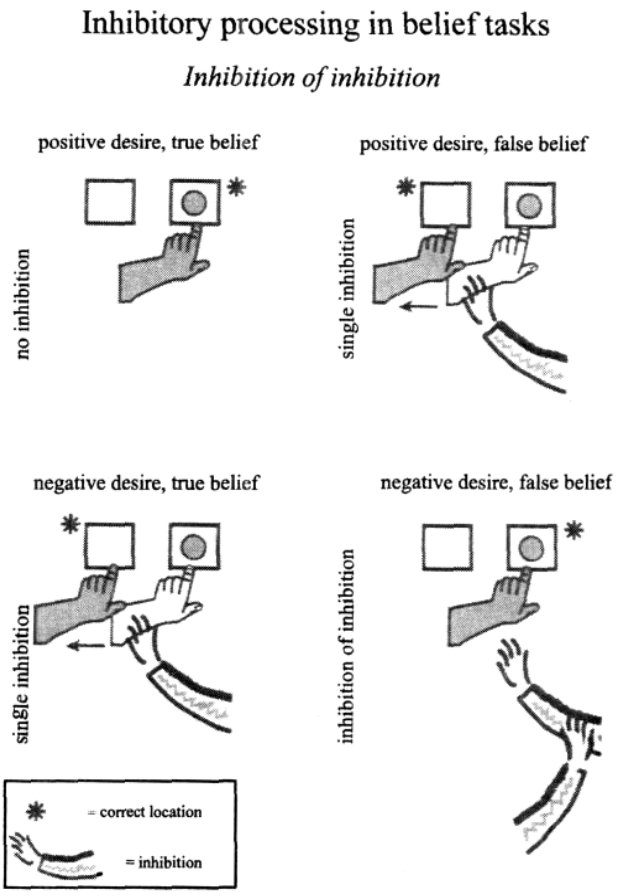

FIGURE 85.8 A model of selection processing in belief-desire reasoning. The panels are arranged to illustrate a 2 X 2 factorial design with rows ± desire and columns ± belief. The pointing hand represents a mental index indicating the target of belief desire and thus, the answer to a "prediction of behavior" question. The grabbing arm represents an inhibitory brain process that weakens an index to which it is applied (reduced shading). Indexes are set initially for true belief and positive desire but subsequently inhibited if the belief is false (second panel) or the desire "negative" (third panel). Weakening of an index causes the index to move to the alternate location. The final panel shows double inhibitions canceling out, rather than summing. This gives the correct answer to false-belief with "negative-desire" problems. (After Leslie and Polizzi, 1998.)

be mobilized, one for the belief because it is false and one for the desire because it is a desire for NOT(X). If both these inhibitions are applied, as before, to the initial target, then it will be inhibited doubly and the index again will move to the empty location. But this time, predicting the empty location is wrong. If Sally wrongly believes the sick kitten to be in the left-hand location and does not want to put the fish in with the kitten, then Sally will try to put the fish in the right-hand location, where the kitten is. To get the correct answer, the two inhibitions cannot be applied in the usual way. Instead, one inhibition must inhibit the other so

that no inhibition reaches the initial target. The index then does not move, and Sally's behavior is predicted correctly. Even though logically, double inhibition problems have the same answer as simple true-belief + positive-desire tasks, the model predicts difficulty from marshaling an inhibition of inhibition.

Leslie and Polizzi (1998) tested the aforementioned prediction on a group of 4-year-olds who passed a standard false-belief task. One half of the children were given a true-belief with target-shifting desire task to measure how difficult it was for them to shift targets from a desire. Only a single child out of 16 tested failed the true-belief task, presumably because shifting from desire targets is easy for 4-year-olds who can target shift in a standard false-belief task. The other half of the subjects were tested with a false belief coupled with target-shifting desire. Here the results were dramatically different. Only 38% of this group correctly predicted which box Sally would approach with the fish.

To answer "think" questions requires calculating belief only; prediction of behavior requires taking into account both belief and desire. Despite this, children's success on "think" and "prediction" in standard tasks is tightly linked. So it is particularly interesting that all of the aforementioned children who failed "prediction" had, immediately before this, passed a "think" question.

The same pattern was found in a second experiment in which Leslie and Polizzi examined whether the aforementioned effects are caused by the linguistic demands of the negation used in stating the desire. Children were introduced to the "Mixed-up Man," a character who always does the opposite of what he wants: if he wants to pat a dog, he pats a cat; if he wants to eat ice cream, he eats a carrot. A scenario was constructed similar to the Sally-with-fish-and-kitten story but with the Mixed-up Man looking for a Mexican jumping bean which, unbeknownst to him, jumps from one box to the other. The key point is that the Mixed-up Man's desire was entirely positive but his behavior was opposite, so that negation did not appear in the protocol. Nevertheless, to predict his behavior, one first had to identify the target of a normal man's action, then shift to the alternate. From the point of view of selection-inhibition theory, what is critical is not negation but whether tasks are processed such that target shifting occurs. Indeed, the false beliefs in standard tasks always are positive but, according to the SP hypothesis, involve target shifting.

The novel Mixed-up Man introduced general difficulty because a substantial number of children failed the true-belief version of the task. Despite being far from ceiling on behavioral target shifting, significantly more children failed the double inhibition version. Results from both experiments are shown in figure 85.9.

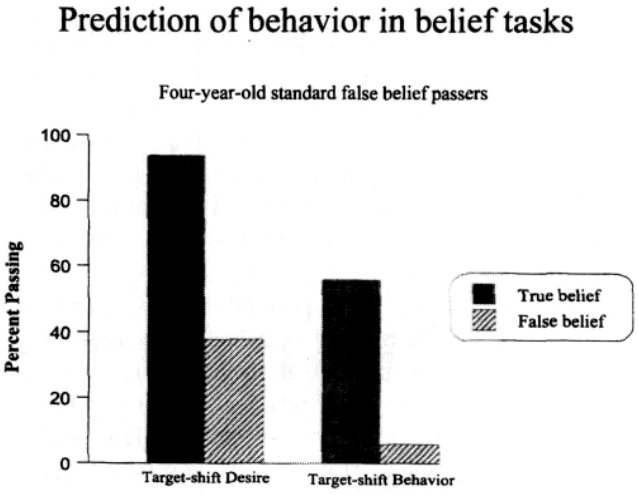

FIGURE 85.9 Four-year-old children who have passed a standard false-belief task can easily combine a "negative" (target-shift) desire with a true belief, but combination with a false belief is very hard. A "target-shift behavior" task was fairly difficult for these children, even with true belief but overwhelmingly difficult when combined with false belief. In both cases, false-belief calculations were difficult even though the false belief was available in memory as shown by ceiling performance on a "think" question. (After Leslie and Polizzi, 1998.)

One interpretation of these results is simply that 4-year-olds are so close to using up all their resources in solving a false-belief task that the addition of *any* further complexity pushes them below threshold. Even the tiny demand from desire-target shifting is sufficient to produce a catastrophic effect. However, this seems unlikely in view of the robustness of 4-year-old performance on a wide variety of standard false-belief tasks (Gopnik 1993), and the lack of reports of other minor modifications that seriously disrupt their performance. However, what is especially intriguing about our findings with double inhibition is that the false-belief calculation should contribute *any* difficulty at all to prediction. Recall that the child solved the false-belief problem to answer the "think" question. All the child has to do then is *remember the answer* for 3 seconds and combine it with *desire* to predict behavior. It is deeply puzzling on a resource model why this task should be any harder than the easy true-belief version: there too, the child simply has to remember the true belief attributed seconds earlier and combine it with desire to predict behavior. It is as if a child is asked to calculate 2 + 2 (hard), manages to get the right answer, then is asked to add 1 to that, and, in response, proceeds to calculate, not 4 +1 (easy), but 2 + 2+1 (extremely hard).

The inhibition of inhibition model explains the aforementioned effect only on the assumption that prediction mandates recalculation of belief despite the answer's availability in memory. Such rigid behavior might reflect the modular character of the ToMM. However, Leslie and Polizzi (1998) proposed a second inhibition model that accounts for these findings without assuming mandatory recalculation. The alternative model is based on the idea of *inhibition of return.* Inhibition of return is an effect familiar from studies of visual attention. It is harder to return attention to a visual target that has been previously attended then disengaged from than it is to attend to the target for the first time (Rafal and Henik, 1994). The model in figure 85.10 assumes that belief (and therefore the belief target) is calculated first and desire targets identified relative to belief. In the critical doubled case, the true-belief target is inhibited first, causing the index to shift to the other location. The desire target is set initially to this second location but, because the desire is "negative," this too must be inhibited, forcing return to the initial location. But the initial location still is inhibited, making return to it difficult. Now, because the first target was inhibited in the process of answering the "think" question correctly, even if the answer is remembered simply for a few seconds, the double inhibition prediction requires return to that still inhibited target. Even without mandatory recalculation, this will be difficult.

One recent study may indicate that recalculation is mandatory. Polizzi and Leslie (1999) tested a group of 4-year-old standard false-belief passers on the "double inhibition" task, outlined previously, but this time instead of asking where Sally would go with the fish, they asked where Sally would go *first.* Recall that asking a "look first" question helps 3-year-old children pass an otherwise standard false-belief task. Surian and Leslie (1999) hypothesized that this was because the word "first" made the nonfactual content more salient, reducing the need for inhibition of the default content. If "look first" works that way for 3-year-olds, might it also help 4-year-olds on double inhibition tasks? Polizzi and Leslie (1999) found that indeed it does: 81% of a group of 4-year-olds succeeded on the double inhibition task when asked the "look first" question. For 4-year-olds to be helped in this way, prediction must force recalculation of belief.

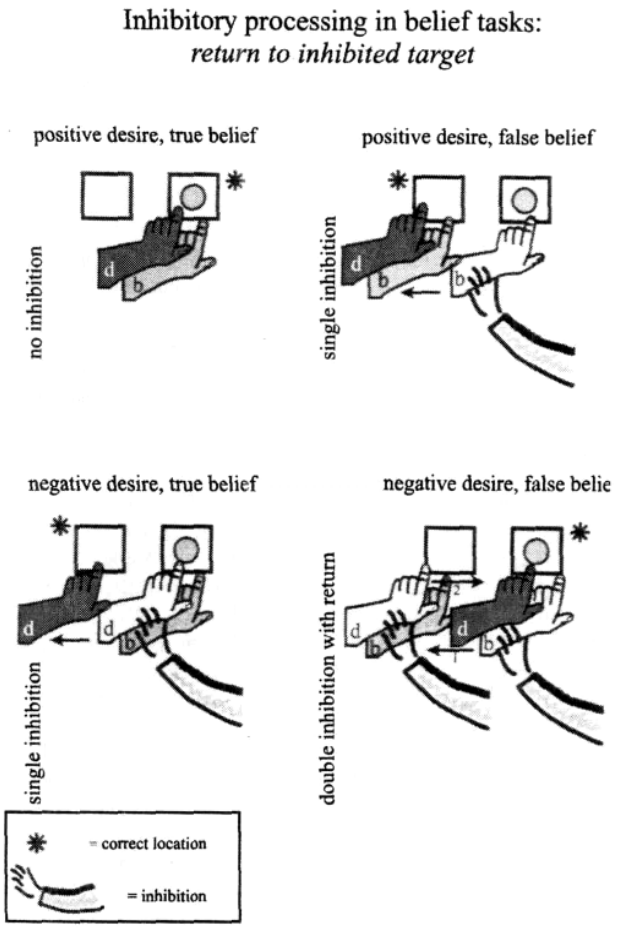

FIGURE 85.10 An alternative model of inhibitory processing in belief-desire reasoning. Instead of identifying the target of belief and desire in parallel, in this model belief targets are identified first and desire targets second. Desire targets are identified in relation to belief targets. Again, indexes are set initially for true belief and positive desire but, subsequently, the belief target is inhibited if the belief is false (second panel) and the desire target inhibited if the desire is "negative" (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 target to move to the alternative. The target of positive desire then is identified in relation to the new false belief target. Finally, the positive-desire target is inhibited, forcing return to the still inhibited true-belief target. (After Leslie and Polizzi, 1998).

Summary

Rather than assume that because mental state concepts are abstract they can only be acquired by the child constructing a theory, I analyze "theory of mind" as a mechanism of selective attention. Mental state concepts simply allow the brain to attend selectively to corresponding mental state properties of agents and thus permit learning about those properties. Autistic children are impaired specifically in this attentional mechanism and find it hard to learn about the mental life of agents. Normally developing preschoolers acquire greater flexibility in attending to the contents of mental states, in particular, to the contents of beliefs that are false. The first information processing models of belief-desire reasoning were outlined.

NOTE

1. Leekam and Perner (1991) found exactly equal numbers of passers and failers, whereas Slaughter (1998) and Perner and associates (in press) used nonparallel task structures, voiding the comparison.

REFERENCES

AMERICAN PSYCHIATRIC ASSOCIATION, 1994. *Diagnostic and Statistical Manual of Mental Disorders,* 4th edition. Washington, D.C.: APA.

BAILEY, A., A. LECOUTEUR, I. GOTTESMAN, P. BOLTON, E. SIMONOFF, E. YUZDA, and M. RUTTER, 1995. Autism as strongly genetic disorder: Evidence from a British twin study. *Psychol. Med.* 25:63-77.

BARON-COHEN, S., 1995. *Mindblindness: An essay on autism and theory of mind.* Cambridge, Mass.: MIT Press.

BARON-COHEN, S., A. M. LESLIE, and U. FRITH, 1985. Does the autistic child have a "theory of mind"? *Cognition* 21:37-46.

CARLSON, S. M., L. J. MOSES, and H. R. HlX, 1998. The role of inhibitory processes in young children's difficulties with deception and false belief. *Child Dev.* 69:672-691.

CASSIDY, K. W., 1995. Use of a desire heuristic in a theory of mind task. Paper presented to the *Biennial Meeting of the Society for Research in Child Development,* April 1995, Indianapolis, Ind.

DIAMOND, A., 1988. Differences between adult and infant cognition: Is the crucial variable presence or absence of language? In *Thought Without Language,* L. Weiskrantz, ed. Oxford: Oxford Science Publications, pp. 335-370.

FRITH, U., J. MORTON, and A. M. LESLIE, 1991. The cognitive basis of a biological disorder: Autism. *Trends Neurosci.* 14: 433-438.

GlLLBERG, C., and M. COLEMAN, 1992. *The Biology of the Autistic Syndromes—2nd Edition. Clinics in Developmental Medicine No. 126.* New York: Cambridge University Press (Mac Keith Press).

GOLDMAN-RAKIC, P. S., 1987. Development of cortical circuitry and cognitive function. *Child Dev.* 58:601-622.

GOPNIK, A., 1993. How we know our minds: The illusion of first-person knowledge of intentionality. *Behav. Brain Sci.* 16: 1-14.

GOPNIK, A., and A. N. MELTZOFF, 1997. *Words, Thoughts, and Theories.* Cambridge, Mass.: MIT Press.

GOPNIK, A., and H. M. WELLMAN, 1995. Why the child's theory of mind really *is* a theory. In *Folk Psychology: The Theory of Mind Debate.* M. Davies and T. Stone, eds. Oxford: Blackwell, pp. 232-258.

HAPPÉ, F. G., 1995. *Autism: An Introduction to Psychological Theory.* Cambridge, Mass.: Harvard University Press.

HARRIS, P. L., and R. KAVANAUGH, 1993. The comprehension of pretense by young children. *Soc. Res. Child Dev. Monogr.* 231.

LEEKAM, S., and J. PERNER, 1991. Does the autistic child have a "metarepresentational" deficit? *Cognition* 40:203-218.

LESLIE, A.M., 1987. Pretense and representation: The origins of "theory of mind." *Psychol. Rev.* 94:412-426.

LESLIE, A. M., 1991. The theory of mind impairment in autism: Evidence for a modular mechanism of development? In *Natural Theories of Mind: Evolution, Development and Simulation of Everyday Mindreading,* A. Whiten, ed. Oxford: Blackwell, pp. 63-78.

LESLIE, A. M., 1992. Autism and the "Theory of Mind" module. *Curr. Dir. Psychol. Sci.* 1:18-21.

LESLIE, A. M., 1994. *Pretending* and *believing:* Issues in the theory of ToMM. *Cognition* 50:211-238.

LESLIE, A. M., and U. FRITH, 1988. Autistic children's understanding of seeing, knowing and believing. *Br.J. Dev. Psychol.* 6:315-324.

LESLIE, A. M., and T. P. GERMAN, 1995. Knowledge and ability in "theory of mind": One-eyed overview of a debate. In *Mental Simulation: Philosophical and Psychological Essays, M.* Davies and T. Stone, eds. Oxford: Blackwell, pp. 123-150.

LESLIE, A. M., and P., POLIZZI, 1998. Inhibitory processing in the false belief task: Two conjectures. *Dev. Sci.* 1:247-258.

LESLIE, A. M., and D. ROTH, 1993. What autism teaches us about metarepresentation. In *Understanding Other Minds: Perspectives from Autism,* S. Baron-Cohen, H. Tager-Flusberg, and D. Cohen, eds. Oxford: Oxford University Press, pp. 83-111.

LESLIE, A. M., and L. THAISS, 1992. Domain specificity in conceptual development: Neuropsychological evidence from autism. *Cognition* 43:225-251.

MITCHELL, P., R. SALTMARSH, and H. RUSSELL, 1997. Overly literal interpretations of speech in autism: Understanding that messages arise from minds. *J. Child Psychol. Psychiatry* 38:685-691.

NAITO, M., S. KOMATSU, and T. FUKE, 1994. Normal and autistic children's understanding of their own and others' false belief: A study from Japan. *Br.J. Dev. Psychol.* 12:403-416.

OZONOFF, S., B. F. PENNINGTON, and S. J. ROGERS, 1991. Executive function deficits in high-functioning autistic individuals: Relationship to theory of mind. *J. Child Psychol. Psychiatry* 32:1081-1105.

PERNER, J., 1991. *Understanding the Representational Mind.* Cambridge, Mass.: MIT Press.

PERNER, J., U. FRITH, A. M. LESLIE, and S. R. LEEKAM, 1989. Exploration of the autistic child's theory of mind: Knowledge, belief and communication. *Child Dev.* 60:689-700.

PERNER, J., S. LEEKAM, D. MYERS, S. DAVIS, and N. ODGERS, in press. Misrepresentation and referential confusion: Children's difficulty with false beliefs and outdated photographs. *Br.J. Dev. Psychol.*

POLIZZI, P. A., and A. M. LESLIE, 1999. "Look first" eases inhibitory demands in the false belief task. Paper presented to the Biennial Meeting of the Society for Research in Child Development, April, Albuquerque, N. Mex.

POSNER, M. I., and D. E. PRESTI, 1987. Selective attention and cognitive control. *Trends Neurosci.* 10:13-17.

PREMACK, D., and G. WOODRUFF, 1978. Does the chimpanzee have a theory of mind? *Behav. Brain Sci.* 4:515-526.

PRIOR, M., B. DAHLSTROM, and T. SQUIRES, 1990. Autistic children's knowledge of thinking and feeling states in other people. *J. Child Psychol. Psychiatry* 31:587-601.

RAFAL, R., and A. HENIK, 1994. The neurology of inhibition: Integrating controlled and automatic processes. In *Inhibitory Processes in Attention, Memory and Language.* D. Dagenbach and T. H. Carr, eds. New York: Academic Press, pp. 1-51.

REED, T, 1994. Performance of autistic and control subjects on three cognitive perspective-taking tasks. *J. Autism Dev. Dis-ord.* 24:53-66.

ROTH, D., and A. M. LESLIE, 1991. The recognition of attitude conveyed by utterance: A study of preschool and autistic children. *Br.J. Dev. Psychol.* 9:315-330.

ROTH, D., and A. M. LESLIE, 1998. Solving belief problems: Toward a task analysis. *Cognition* 66:1-31.

SIEGAL, M., and K. BEATTIE, 1991. Where to look first for children's knowledge of false beliefs. *Cognition* 38:1-12.
SLAUGHTER, V., 1998. Children's understanding of pictorial and mental representations. *Child Dev.* 69:321-332.
SODIAN, B., and U. FRITH, 1992. Deception and sabotage in autistic, retarded and normal children. *J. Child Psychol. Psychiatry* 33:591-605.
SURIAN, L., and A. M. LESLIE, 1999. Competence and performance in false belief understanding: A comparison of autistic and three-year-old children. *Br.J. Dev. Psychol.* 17:141-155.
TAGER-FLUSBERG, H., 1992. Autistic children's talk about psychological states: Deficits in the early acquisition of a theory of mind. *Child Dev.* 63:161-172.
WIMMER, H., and J. PERNER, 1983. Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition* 13: 103-128.
WING, L., and J. GOULD, 1979. Severe impairments of social interaction and associated abnormalities in children: Epidemiology and classification. *J. Autism Dev. Disord.* 9:11-29.
ZAITCHIK, D., 1990. When representations conflict with reality: The preschooler's problem with false beliefs and "false" photographs. *Cognition* 35:41-68.