RuCCS TR–10

Pretending and Believing

Issues in the theory of ToMM

Alan M. Leslie
Center for Cognitive Science
Department of Psychology
Rutgers University
aleslie@ruccs.rutgers.edu

Dedication: This article is dedicated to the memory of Daniel Roth, my student, collaborator and friend who tragically lost his long struggle against cancer on 17th April 1993.

Acknowledgements: This paper has undergone a long gestation, various parts having been presented to the BPS Developmental Section Annual Conference, Coleg Harlech, September 1988, *Trieste Encounters on Cognitive Science*, Trieste, Italy, May 1989, *International Workshop on Naturalized Epistemology*, Cornell University, December 1989, *International Conference on Cultural Knowledge and Domain Specificity*, University of Michigan, Ann Arbor, October 1990, *Society for Research on Child Development Biennial Meeting*, Seattle, April 1991, and *Inaugural Conference of the Rutgers University Center for Cognitive Science*, Rutgers University, November 1991. I am grateful to participants and audiences at those meetings and also to colleagues and friends at the MRC Cognitive Development Unit for nurture and good nature.

Technical Report #10, November 1993
Rutgers University Center for Cognitive Science
Rutgers University
PO Box 1179
Piscataway, NJ 08855

ABSTRACT

Commonsense notions of psychological causality emerge early and spontaneously in the child. What implications does this have for our understanding of the mind/brain and its development? In the light of available evidence, the child's "theory of mind" is plausibly the result of the growth and functioning of a specialized mechanism (**ToMM**) *that produces domain-specific learning. The failure of early spontaneous development of "theory of mind" in childhood autism can be understood in terms of an impairment in the growth and functioning of this mechanism.* **ToMM** *constructs Agent-centered descriptions of situations or "metarepresentations." Agent-centered descriptions place Agents in relation to* information. *By relating behaviour to the attitudes Agents take to the truth of propositions,* **ToMM** *makes possible a commonsense causal interpretation of Agents' behaviour as a result of circumstances that are imaginary rather than physical. Two early attitude concepts,* pretends *and* believes, *are discussed in the light of some current findings.*

Consider the scenario in Figure 1. Numerous studies (e.g. Baron-Cohen, Leslie & Frith, 1985; Wimmer & Perner, 1983) have shown that, by about 4 years, children understand this scenario by attributing a (false) *belief* to Sally and predicting her behaviour accordingly. Premack and Woodruff (1978) coined the term "theory of mind" for the ability, illustrated by this scenario, to predict, explain and interpret the behaviour of Agents in terms of mental states. Such findings raise the following question. How is the preschool child able to learn about mental states when these are unobservable, theoretical constructs? Or put another way: How is the young brain able to attend to mental states when they can be neither seen, heard nor felt?

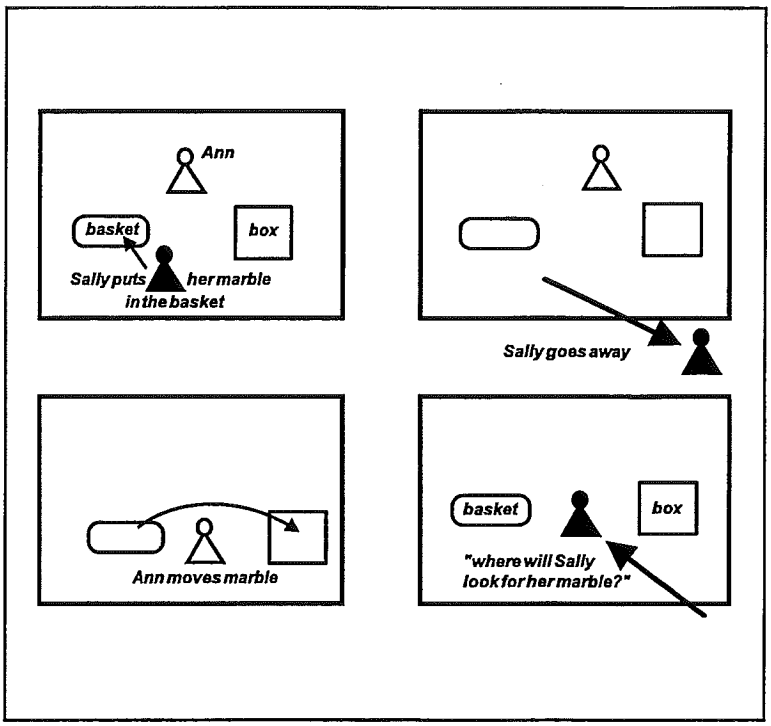

Figure 1. A standard false belief (FB) scenario that can be solved by four-year-olds (after Baron-Cohen, Leslie & Frith, 1985).

A general answer to the above question is that the brain attends to behaviour and infers the mental state that the behaviour issues from. For example, in the scenario in Figure 2, Mother's behaviour is *talking to a banana*. The task for a two year old watching her is to infer that **Mother PRETENDS (of) the banana (that) "it is a telephone"**. Mother's behaviour described as a physical event—as one object in relation to another—is minimally interesting. The real significance of her behaviour emerges only when mother is described as an Agent in relation to information. As an Agent, mother can adopt an attitude (of pretending) to the truth of a description ("it is a telephone") in regard to a given object (the banana). Entertaining this kind of intentional or Agent-centered description requires computing a certain kind of internal representation. I have called this the "metarepresentation" or "M-representation" (Leslie, 1987; Leslie & Thaiss, 1992).

I shall explore the following assumption. Native to our mental architecture is a domain specific processing stream adapted for understanding the behaviour of Agents. A major component of this system is a mechanism which computes the M-representation. I call this mechanism **ToMM** (theory of mind mechanism). Here are five guiding ideas in the theory of **ToMM**:

1) *The key to understanding the origins of theory of mind lies in time-pressured, on-line processing to interpret an Agent's behaviour in terms of underlying intentions.* Early in development, human beings undertake the information processing task of understanding the behaviour of Agents, not simply as a sequence of events, but as instantiating intentions in the broad sense, that is, as issuing from mental states. This processing task is time-pressured because Agent-centered descriptions must be arrived at fast enough to keep up with the flow of behaviour in a conversation or other interaction. This pressure will constrain the amount and types of information that can be taken into account and has had an adaptive evolutionary influence on the architecture of theory of mind processing.

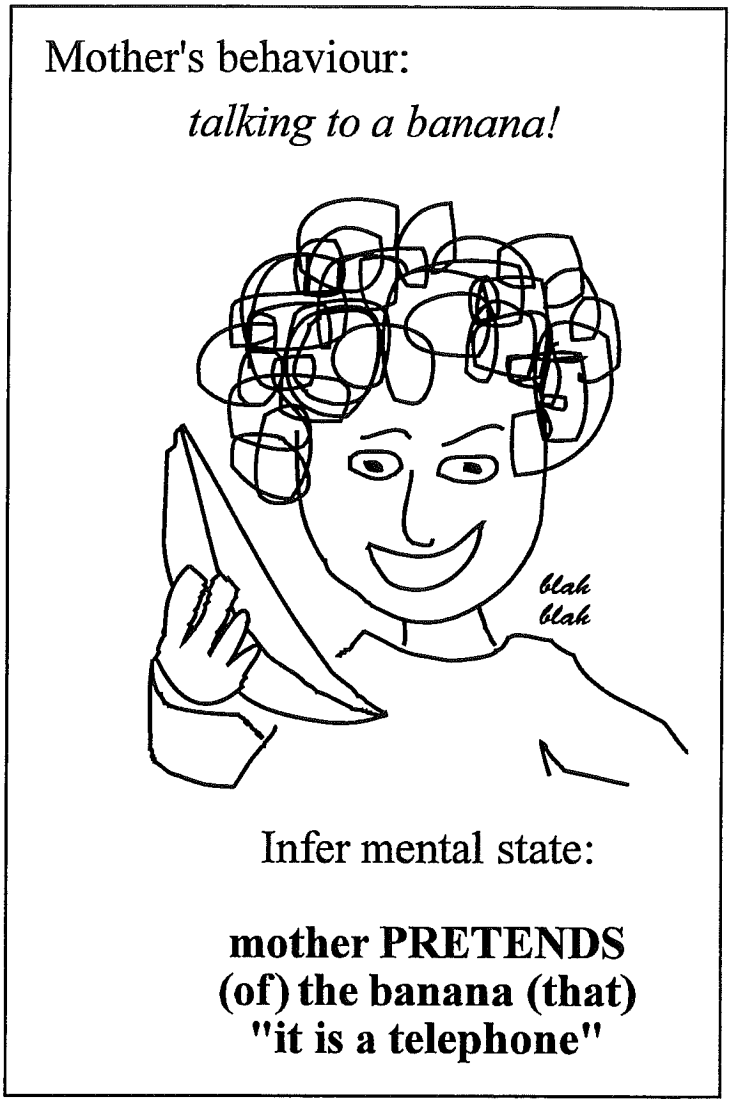

Figure 2. A pretend scenario that can be solved by two-year-olds.

2) *Descriptions of intentional states are computed by a specialized theory of mind mechanism (**ToMM**) which is post-perceptual, operates spontaneously, is domain specific, and is subject to dissociable damage—in the limit, modular.* Information about behaviour arrives through a number of different sensory channels and includes verbal utterances, so **ToMM** should operate post-perceptually. **ToMM** should be able to function relatively spontaneously since it has the job of directing the child's attention to mental states which, unlike behaviour, cannot be seen, heard or felt. **ToMM** should also be able to function as a source of intuitions in reasoning about Agents and thus be addressable centrally. **ToMM** is specifically concerned with "cognitive" properties of Agents and employs specialized notions for this task. Finally, **ToMM** can be damaged or impaired independently of other processing systems (see below).

3) **ToMM** *employs a proprietary representational system which describes propositional attitudes.* This property of **ToMM** is discussed in the theory of the M-representation to which I return below.

4) **ToMM** *forms the specific innate basis for our capacity to acquire a theory of mind.* Perhaps the most important job **ToMM** has to do is to produce development within its designated domain and to produce it early, rapidly and uniformly without benefit of formal instruction. To this end, **ToMM** introduces the basic *attitude* concepts and provides intuitive insight into mental states early in life while encyclopedic knowledge and general problem solving resources are limited.

5) **ToMM** *is damaged in childhood autism resulting in its core symptoms and impairing these children's capacity to acquire a theory of mind.* Leslie and Roth (1993) have recently reviewed evidence supporting this idea (see also Frith, Morton & Leslie, 1992; Leslie & Frith, 1990).

Pretending and **ToMM**

One of the earliest easily observed products of **ToMM** the capacity to pretend. The ability to pretend emerges between 18 and 24 months of age. Around this time, the child begins to entertain deliberate suppositions about simple imaginary situations: for example, she pretends that a banana is a telephone or that an empty cup contains juice. The ability is productive and does not remain limited to a single or to a few special topics, is exercised playfully and communicatively without ulterior motive (e.g. to deceive), permits sharing of information about imaginary situations with other people, and encompasses the ability to understand other people's communicative pretence. Due regard must be paid to the question of distinguishing pretence from other phenomena which are superficially similar at a behavioral level (e.g. functional play, acting from a mistaken belief, play in animals). I discussed some of the more important of these distinctions in Leslie (1987) and pointed out that the aim of previous workers to develop a *behavioural* definition of pretence was unattainable. I proposed instead a theoretical definition in terms of underlying cognitive processes.

There are four critical features of early pretence that a cognitive model must capture. The first requirement is to account for the fundamental forms of pretence. There are three of these, one for each of the basic (external) semantic relations between a representation and what it represents (viz., reference, truth and existence). In object substitution pretence, a given real object, e.g. a banana, is pretended to be some other object, e.g. a telephone. Such pretence requires a decoupling of the internal representation for telephones from its normal *reference* so that it functions in context as if it referred to a member of some arbitrary class of object, in this case a banana. Second, in properties pretend, a given object or situation is pretended to have some property typically it does not have, e.g., a dry table is pretended to be wet. Here the pretence decouples the normal effects of *predicating* wetness in the internal representation. And thirdly, imaginary objects can be pretended to have *existence*, e.g., that there

is a hat on teddy's head. Here the pretence affects the normal existence presuppositions in the internal representation. A cognitive model of pretence has to explain why there are exactly three fundamental forms and why there are exactly these three.[1]

Leslie (1987) argued that the fundamental forms of pretence reflect the semantic phenomena of *opacity* (Quine, 1961). Opacity may be roughly described as the result of the "suspension" of the semantic relations of reference, truth and existence that occurs when a representation is placed in an intentional contexts, such as a mental state report or counterfactual reasoning. To explain the isomorphism between the three fundamental forms of pretence (behavioural phenomena) and the three aspects of opacity (semantic phenomena), I proposed the existence of a certain kind of internal representation. Representational structures, having whatever properties give rise to the opacity phenomena of pretence, must be a property of the human mind/brain from its infancy onwards.

The second critical feature of the development of pretence that a cognitive theory must account for is related to the first. Rather than appearing in three discrete stages, the fundamental forms of pretence emerge together in a package. Given a *single* mechanism with the right properties, a cognitive model can capture both the fact of the three fundamental forms of pretence and their emergence as a package (see Leslie, 1987 for discussion).

The third crucial feature of pretence to be explained is, when the child first becomes able to pretend herself (solitary pretence), why does she also gain the ability to understand pretence-in-others? Traditional investigations overlooked this startling fact. Understanding another person's behaviour as a pretence can be studied as an information processing task the child undertakes. For example, when mother says, "The telephone is ringing", and hands the child a banana, the two year old, who is undertaking a number of complex information processing tasks simultaneously, such as, building a catalogue of object kinds, analysing Agents' goal directed actions with instruments and acquiring a lexicon, is in general neither confused about bananas, nor about mother's strange behaviour, nor about the meaning of the word "telephone". Instead, in general, the child understands that mother's behaviour—her gesturing and her use of language—relates to an imaginary situation which mother pretends is real. Again, we can account for the yoking in development between the capacity to pretend oneself and the capacity to understand pretence-in-others if we assume that a single mechanism is responsible for both.

---

[1] For reasons which are not clear, Perner (1991) writes as if the fact that the fundamental forms can be combined into more complex forms—for example, pretending that teddy's imaginary hat has a hole in it—should be a source of embarrassment to my theory. In fact, the possibility of "complex" pretence springs readily from the assumed combinatorial properties of metarepresentation. A further misunderstanding is to suppose that the only way the child could possibly handle the three aspects of opacity is by explicitly theorizing about reference, truth and existence—i.e., by theorizing about the general nature of representation. Leslie (1987) did not propose any such thing for understanding pretence. Indeed the whole thrust of my proposals was to avoid such a commitment by describing processing machinery that would achieve a similar result implicitly.

Finally, a cognitive account must address the fact that pretence is related to particular aspects of the *here and now* in specific ways. This is true both for solitary pretence and in understanding the pretence of other people. For example, it is *this* banana that mother pretends is a telephone, not bananas in general nor *that* banana over there. The truth of the pretend content, **"it is a telephone"**, is anchored in a particular individual object in the here and now. This is another critical feature of the early capacity for pretence that a cognitive model must capture.

These four critical features of pretence—the three fundamental forms, their emergence as a package, the yoking of solitary pretence with the ability to understand pretence-in-others, and the anchoring of pretend content in the here and now—can be succinctly explained as consequences of the data structure called the "metarepresentation". This representational system provides precisely the framework that is needed to deploy another attitude concept closely related to *pretending*, namely, the concept of *believing*. Thus, the same representational system is required if the child is to interpret mother's behaviour in terms of **mother** BELIEVES **(of) the banana (that) "it is a telephone"**. *Pretending* and *believing*, though closely related attitude concepts, are, nevertheless, different concepts and their successful deployment can make rather different demands on problem solving resources. I shall consider the emergence of the concept, *believing*, in the second part of this article.

The Metarepresentation

Leslie (1987, 1988c; Leslie & Frith, 1990) outlined some general ideas on how a mechanism like **ToMM** could achieve the above solution. Three different types of representation were distinguished. "Primary" representations are literal, transparent descriptions of the world. "Decoupled" representations are opaque versions of primary representations. The decoupling of a representation allows a processor to treat the representation as a "report" of information instead of merely being reacted to it. This in turn allows the (decoupled) representation to be placed within a larger relational structure in which an attitude to the truth of the "report" can be represented. This larger relational structure is built around a set of primitive relations—the *attitude* concepts or "Informational Relations". These relations tie together the other components. This entire relational structure is the third type of representation and is referred to as the "metarepresentation" (or, to distinguish it from Perner's later use of the term, the "M-representation").

ToMM employs the system of metarepresentation. Following Marr (1982), we can say that this system makes explicit four kinds of information. Descriptions in this system identify:

1) an Agent
2) an Informational Relation (the attitude)
3) an aspect of the real situation (the anchor)
4) an "imaginary" situation (the description),

such that a given Agent takes a given attitude to the truth of a given description in relation to a given anchor. The Informational Relation is the pivotal piece of

information in the sense that it ties together the other three pieces of information into a relational structure and identifies the Agent's attitude. The direct object of the identified attitude is (the truth of) a proposition or description (typically of an "imaginary" situation) in relation to a "real" object or state affairs. Not counting the implicit truth value, Informational Relations are thus 3-place relations (Leslie, 1987).

What does an Informational Relation represent? Perner (1991) has made a great deal of the fact that I borrowed the term "metarepresentation" from Pylyshyn (1978) for whom it meant a "representation of the representational relation". This seemed an innocuous enough phrase to me then, and still does, as long as one leaves it as an *empirical* issue exactly how a given "representational relation" is represented. But for Perner the term can only mean that the child possesses a certain kind of "representational theory of mind" (RTM) in which mental states are individuated by form rather than by meaning. I see no reason to accept this stricture. In any case, Leslie (1987) simply assumed that very young children did *not* have access to a RTM in this sense. The model of metarepresentation I outlined was designed to account for the very young child's capacities by attributing more modest knowledge in which, for example, "representational relations", such as reference and truth, are handled implicitly, while "representational relations" such as *pretending* and *believing* are handled explicitly. As we shall see later, there is no evidence available to suggest that preschool children have a RTM in Perner's sense.

The critical point about what Informational Relations represent is that they denote the kind of relation that can hold between an *agent* and the truth of a *description* (applied to a given state of affairs). This immediately determines a class of notion different from other kinds of relation that feature in early cognition, for example, spatial and mechanical relations, and forms the conceptual core of commonsense theory of mind.

My assumption is that there is a small set of primitive Informational Relations available early on, among them BELIEVE and PRETEND. These notions are primitive in the sense that they cannot be analyzed into more basic components such that the original notion is eliminated. While one can paraphrase "John believes that *p* is true" in a number of ways, one does not thereby eliminate the notion *believes*. For example, one can say "p is true for John", but that just gives another way (an alternate set of sounds for) saying "John believes that *p* is true".

Perner (1991) adopts part of the above theory of pretence, namely the notion of decoupling, though he discusses it in terms of "models". According to this view, pretence emerges when the child can entertain multiple "models" of the world instead of just the single model that is possible during the first year. Representations of different times and places apparently constitute different models. It seems unlikely that infants during the first year cannot relate past states of affairs to present ones but, in any case, Perner's notion of "model" is does not say much about pretence. The opacity properties of pretence are not illuminated by tense and location "models" because the content of pretence is opaque *in the here and now*. This kind of opacity is also what is relevant to *believing*. By contrast, in a "Zaitchik photograph" (one that has gone out of date), the photograph is only a representation of a past situation and not of the

current situation. Compare this case to the case in which someone assumes (wrongly) that the photograph is a photograph of the current situation. This is quite a different matter. The critical feature in this latter case is clearly not the representation itself which remains the same, but the fact that an Agent *believes* that the photograph depicts a current situation.

What Perner's model notion fails to address is the relationship of the agent to the "model". Perner (1988) says that for the child the agent is simply "associated" with the model, though an associative relationship can also hold between, for example, can-openers and kitchens without the child ever thinking that can-openers pretend anything about kitchens. Perner (1991) at times seems inclined to attribute a behaviourist notion of pretence to the young child such that the agent who pretends that *p* is understood as *acting as if p* were true. This proposal is only useful if we are also told how the child views the relation between *p* and the agent's behaviour in the case in which *p* actually is true. If the relation between circumstances and behaviour in the normal case is causal, is it also causal in the case of pretence? If so, how can imaginary circumstances be viewed as causal? How could the child learn about the causal powers of imaginary circumstances? The only solution to this dilemma, as far as I can see, involves some kind of mentalistic rather than behaviouristic interpretation of the relation between the agent and *p*, that is, some kind of *attitude* notion. Finally, parity of argument demands that if we insist upon a behaviouristic construal of pretence-understanding in the child, then we should also insist upon a behaviouristic construal of false-belief-understanding. After all, falsely believing that *p* demands the interpretation *acting as if p* every bit as much (or every bit as little, depending upon point of view) as pretending that *p*. The fact that one child is a bit older than the other does not in itself constitute a compelling reason for treating the two cases in radically different ways.

Decoupling

The role of decoupling in the metarepresentation is to transform a primary, transparent internal representation into something that can function as the direct object of an Informational Relation. In the case of Informational Relations, as in the case of verbs of argument and attitude, the truth of the whole expression is not dependent upon the truth of its parts. This is a crucial part of the semantics of mental state notions and what gives rise to the possibility of pretends and beliefs being false. The decoupling theory was an attempt to account for this feature without supposing that the *child* had to devise a theory of opacity. Leslie (1987) suggested that one way to think about the decoupling of an internal representation from its normal input/output relations vis a vis normal processing was as a report or copy in one processing system (the "Expression Raiser") of a primary representation in another (e.g. general cognitive systems). This suggestion drew upon the analogy between opacity phenomena in mental state reports and reported speech. Subsequently, Leslie (1988c) and Leslie and Frith (1990) developed this idea in terms of the relationship between decoupled representations and processes of inference. The basic idea is that decoupling introduces extra structure into a representation and that this extra structure affects how processes of inference operate, ensuring that the truth of the part does not determine the truth of the whole. The

simplest illustration of this is that one cannot infer **it is a telephone** from **"it is a telephone"**.

Normally, the truth of a whole expression is determined by the truth of its parts. For example, "Mary picked up the cup which was full" is true *iff* the cup Mary picked up was full. This same principle is involved in the detection of contradiction. Consider the following: **the cup is empty** and **the empty cup is full**. Suppose the whole-parts principle was implemented in a spontaneous inferencing device that carries out elementary deductions. The device will quickly produce the conclusion, "the cup is full & not full", revealing a contradiction because this whole and all of its parts cannot be simultaneously true. Despite the surface similarity to the foregoing, however, the device should not detect a contradiction in **I pretend the empty cup is full**. One might think at first that contradiction is blocked by the element, **pretend**, but contradiction returns in **I pretend the cup is both empty and full**, despite the presence of the element, **pretend**. We can think of decoupling as controlling the occurrence of contradiction:

(1) **the cup is empty**
(2) **the empty cup is full**
(3) **I pretend the empty cup "it is full"**
(4) **I pretend the cup "it is both empty and full"**.

Decoupling creates extra structure in the representation—an extra level to which the inferencing device is sensitive. Thus, in (2), with no decoupling, there is a single level within which a contradiction is detectable. In (3), there are two levels. The inferencing device first examines the upper level where it encounters **I pretend the empty cup X** and registers no contradiction. Next, it examines the lower level where it sees **X "it is a telephone"** and again detects no contradiction. On the lower level of (4), however, the device encounters **X "it is both empty and full"** and registers contradiction within the level as in (2). Contradiction is detected within but not across decoupled levels. This is exactly what is required by informational relations.

Similar patterns can be seen in causal inferences. For example, **I pretend this empty cup "it contains tea"** can be elaborated by an inference such as: *if a container filled with liquid is UPTURNED, then the liquid will pour out and make something wet.* This same inference works in both real and pretend situations; it also works for both own pretence and for understanding other people's pretence (Leslie, 1987). Of course, in pretend situations, we do not conclude that pretend tea will really make the table wet. The consequent is decoupled because the antecedent was. So, if I upturn the cup which I am pretending contains tea, I conclude that **I pretend the table "it is wet"**. The conclusions of the inference are again closed under decoupling; or we may say that the inference operates within the decoupled level.

If pretend scenarios—both one's own and those one attributes to other people—unfold by means of inference, then we could predict another consequent based on a variation of the above inference: *if the liquid comes out of the container, then the container will be empty.* This leads to pretending something that is true, namely, that the empty cup is empty. At first glance, this may seem ridiculous. But there is, of course, an important difference between **the empty cup is empty** and **pretending (of)**

the empty cup "it is empty". Later I will present an empirical demonstration that young children routinely make this sort of inference in pretence.

Yoking. The emergence of solitary pretence is yoked to the emergence of the ability to understand pretence-in-others. The very young child can share with other people the pretend situations she creates herself and can comprehend the pretend situations created by other people. She is able to comprehend the behaviour and the goals of other people not just in relation to the actual state of affairs she perceives, but also in relation to the imaginary situation communicated to her and which she must infer. We can illustrate this in two different ways: first in relation to behaviour, and second in relation to language use. Mother's goal-directed behaviour with objects will be an important source of information for the young child about the conventional functions of objects. Likewise, mother's use of language will be a major source of information about the meanings of the lexical items the child learns. But in pretence, the child will have to know how to interpret mother's actions and utterances with respect to mother's pretence rather than with respect to the primary facts of the situation. When mother says, e.g., "The telephone is ringing" and hands the child the banana, it will not be enough for the child to compute linguistic meaning. She will have to calculate *speaker's meaning* as well as (cf. Grice, 1957). This double computation is inherently tied to the Agent as the source of the communication and is seamlessly accomplished through the metarepresentation.

Interestingly, Baldwin (in press) has provided independent evidence that children from around 18 months of age begin to calculate speaker's meaning. In the circumstances studied by Baldwin, the 18 month old calculates speaker's meaning, not in service of pretence, but in the service of calculating linguistic meaning. Baldwin showed that, from around 18 months, children do not simply take the utterance of a novel word to refer to the object they themselves are looking at but instead look round and check the gaze of the speaker. They then take the novel word to refer to the object that the speaker is looking at, even if this is different from the one they were looking at when they heard the utterance. This finding reinforces the idea that infants around this age are developing an interest in what might be called the "informational properties" of Agents.

Understanding pretence-in-others

In this section, I shall describe an experimental demonstration of a number of the phenomena discussed so far. This experiment was first presented in Leslie (1988a) and discussed briefly in Leslie (1988c).[2] The following hypotheses are tested. First, early pretence can involve counterfactual inferencing. Second, such inferencing can be used to elaborate pretence. Third, inferencing within pretence can use real world causal

[2] Harris and Kavanaugh (in press) have recently replicated and extended this study, though they draw somewhat different conclusions in line with their "simulation theory". The simulationist view of theory of mind phenomena raises a number of complex issues which I do not discuss here (but see Leslie and German, in press).

knowledge and that such knowledge is available to two-year-olds in a form abstract enough to apply in imaginary situations and in counterfactual argument where perceptual support is minimal or contradictory. Fourth, two-year-olds can infer the content of someone else's pretence and demonstrate this by making an inference appropriate to that person's pretence. Fifth, two-year-old pretence is anchored in the here and now in specific ways. Sixth, that pretend contents are not always counterfactual. Seventh, that one can communicate through action, gesture and utterance a definite pretend content to a two-year-old child, sufficient for the child to calculate speaker's meaning/pretender's meaning and to support a particular counterfactual inference based upon the communicated content.

Method

The child was engaged by the experimenter in pretend play. Toy animals and some other props were introduced to the child during a warm-up period. My assistants in this task were Sammy Seal, Mummy Bear, Lofty the Giraffe, Larry Lamb and Porky Pig. Other props included toy cups, plates, a bottle, some wooded bricks and a paper tissue. The experimenter pretended that it was Sammy's birthday that day, that Sammy was being awakened by Mummy Bear and was being told that there was going to be a party to which his friends were coming. This warm up period served to convey that what was to happen was pretend play and to overcome the shyness children of this age often and quite rightly have with strangers who want to share their innermost thoughts with them.

The general design was to share pretence, allowing the child to introduce what elements he or she wished or felt bold enough to advance but to embed a number of critical test sub-plots as naturally as possible into the flow of play. These sub-plots allow testing of pretence appropriate inferencing. Could the child make inferences which are appropriate to the pretend scenario he has internally represented, but which are not appropriate to the actual physical condition of the props?

The sub-plots.

1). CUP EMPTY/FULL. The child is encouraged to "fill" two toy cups with "juice" or "tea" or whatever the child designated the pretend contents of the bottle to be. The experimenter then says, "Watch this!", picks up one of the cups, turns it upside down, shakes it for a second, then replaces it alongside the other cup. The child is then asked to point at the "full cup" and at the "empty cup". (Both cups are, of course, really empty throughout.)

2) UPTURN CUP. Experimenter "fills" a cup from the bottle and says, "Watch what happens!". Sammy Seal then picks up the cup and upturns it over Porky Pig's head, holding it there upside down. Experimenter asks, "What has happened? What has happened to Porky?"

3) MUDDY PUDDLE. The child is told that it is time for the animals to go outside to play. An area of the table top is designated "outside". A sub-part of this area is pointed to

and Experimenter says "Look, there's a muddy puddle here!". Experimenter then takes Larry Lamb and says "Watch what happens". Larry is then made to walk along until the "puddle" area is reached whereupon he is rolled over and over upon this area. Experimenter asks "What has happened? What has happened to Larry?"

4) BATH-WATER SCOOP. Following the above, it is suggested that Larry should have a bath. Experimenter constructs a "bath" out of four toy bricks forming a cavity. Experimenter says, "I will take off Larry's clothes and give him a bath. Then it will be your turn to put his clothes back on. OK?" Experimenter then makes movements around the body and legs of Larry suggesting perhaps the removal of clothes and each time puts them down on the same part of the table top making a "pile".

Larry is then placed in the cavity formed by the bricks for a few seconds while finger movements are made over him. Larry is then removed and placed on the table. Experimenter then says, "Watch this!" and picks up a cup. The cup is placed into the cavity and a single scooping movement is made. The cup is then held out to the child and he or she is asked, "What's in here?" If the child does not answer, the scoop is repeated once to "Watch this" and "What's in here?".

5) CLOTHES PLACE. The child is told "It's your turn to put Larry's clothes on again" and handed Larry. Where (if anywhere) the child reaches in order to get the "clothes" is noted.

Subjects. There were 10 children aged between 26 and 36 months with a mean age of 32.6 months. Two further children were eliminated for being uncooperative or wholly inattentive.

Results

Table 1 shows the number of children passing each sub-plot plus the entire range of responses that occurred. The failures came from 2 children who answered "Don't know" or failed to respond despite the sub-plot being repeated for them.

Statistical analysis seems mostly unnecessary. The CUP EMPTY/FULL sub-plot could be guessed correctly half the time, so all ten children passing is significant ($p = 0.001$, Binomial Test). In the other cases it is difficult to estimate the probability of a correct answer by chance but it is presumably low.

Discussion

These results support a number of features of the metarepresentational model of pretence. They demonstrate counterfactual causal reasoning in two year olds based on imaginary suppositions. For example, in the Cups Empty/Full scenario the child works from the supposition **the empty cups "they contain juice"** and upon seeing the experimenter upturn one of the cups, the child applies a "real world" inference concerning the upturning of cups (see page ?). In this case the child was asked about the cups, so the conclusions generated were, **this empty cup "it contains juice"**, and, **that empty cup "it is empty"**. The last conclusion is, of course, an example of

Table 1. Number of subjects passing test sub-plots and the full range of responses obtained.

Test	*Subjects passing*	*Range of responses obtained indicating appropriate inference.*
CUP EMPTY/FULL cup	10/10	points to or picks up correct
UPTURN CUP	9/10	refills cup, says "I'll wipe it off him" and wipes with tissue, "threw water on him", "he's spilling", "he got wet", "poured milk over him...wet", "tipped juice on head".
MUDDY PUDDLE	9/10	dries animal with tissue, says "oh no, all the mud", "covered in mud", "got mud on".
BATH-WATER SCOOP	9/10	says "water", "water" and pours into other cup, "water" and upturns into bath, "bath-water".
CLOTHES-PLACE	9/9	picks up from correct place, points to correct place.

Failures were produced by two different children with "don't know" responses or no response after the test was repeated

pretending something which is true and not counterfactual. Notice however that in terms of decoupling this is not the tautology, **this empty cup is empty**. A similar conclusion was generated by one of the children in the UPTURN CUP scenario and expressed by him pretending to refill the **"empty cup"** when asked what had happened. These examples help us realise that, far from being unusual and esoteric, cases of "non-counterfactual pretence", i.e. pretending something is true when it is true, are ubiquitous in young children's pretence and indeed has an indispensable role in the child's ability to elaborate pretend scenarios. This is predicted by the Leslie (1987) model.[3]

One way to understand the above result is this: the *logic* of the concept, pretend, does not require that its direct object (i.e., its propositional content) be false. Our feeling that a "true pretend" is odd reflects the *normativity* of our concept of pretence. Having counterfactual contents is, as it were, what pretends are *for*; pretends "ought" to be false; but their falseness is not strictly required by the logic of the concept. In this regard, PRETEND shows the logic of the BELIEVE class of attitudes: the truth of the whole attitude expression is not dependent upon the truth of all of its parts, specifically, not a function of the truth of its direct object. As we saw earlier, we can understand this peculiarity of attitude expressions in terms of decoupling. Some attitudes, like KNOW, on the other hand, do require the truth of their direct object (though even here there are subtleties), but PRETEND and BELIEVE do not. As we shall see later, BELIEVE shows the opposite normativity to PRETEND. Normally, beliefs are true.

In the experiment, the children correctly inferred what the experimenter was pretending. The very possibility of "correctness" depends upon some definite pretend situation being communicated. The child calculates a construal of the Agent's behaviour; a construal which relates the Agent to the imaginary situation that the Agent communicates. The child is not simply socially excited into producing otherwise solitary pretend; the child can answer questions by making definite inferences about a definite imaginary situation communicated to him by the behaviour of the Agent. To achieve this, the child is required to calculate, in regard to utterances, speaker's meaning as well as linguistic meaning, and, in regard to action, the Agent's pretend goals and pretend assumptions. The child is also capable of intentionally communicating his own pretend ideas back to participating Agents.

One of the deep properties that we seem pre-adapted to attribute to Agents is the power of the Agent to take an attitude to imaginary situations (or more accurately, to the truth of descriptions). This allows a rational construal of the role of non-existent affairs in the causation of real behaviour. It is striking that this is done quite intuitively by very young children. The spontaneous processing of the Agent's utterances, gestures and mechanical interactions with various physical objects to produce an interpretation of Agent pretending this or that is surely one of the infant's more sublime accomplishments. However, there is no more need to regard the child as "theorising" like a

---

[3] Though I was perhaps the first to derive this as a prediction from a theoretical model, I am certainly not the first to point out that pretends "can be true". Vygotsky (1967) describes the case of two sisters whose favourite game was to pretend to be sisters.

scientist when he does this than there is when the child acquires the grammatical structure of his language.

Believing and **ToMM**

One of the central problems in understanding the development of theory of mind is the relation between the concepts of *pretending* and *believing.* There are two broad possibilities. There may be no specific relationship between the two and their development may reflect quite different cognitive mechanisms and quite different representational structures. Versions of this position have been held for example by Perner (1991), Flavell (1988), Gopnik & Slaughter (1991).

Alternatively, there may be a close psychological relationship between the concepts of *pretending* and *believing*: both may be introduced by the same cognitive mechanism; both may belong to the same pre-structured representational system. Within this general scheme there are a number of more detailed options. For example, one concept, *pretend*, may develop first while the other, *believe*, may develop later, either because of maturational factors or because *believe* requires more difficult and less accessible information to spur its emergence (Leslie, 1988b). Or the two notions may differentiate out of a common ancestor concept. Or there may be a progressive strengthening or sharpening of the pre-structured metarepresentational system (Plaut and Karmiloff-Smith, 1993). Or there may be different performance demands associated with employing the two concepts—demands that can be met only at different times in development depending upon a variety of factors (Leslie & Thaiss, 1992). Whichever of these options, or whichever mixture of these options (they are not mutually exclusive), turns out to be correct, there is no reason to suppose that *pretend* and *believe* require radically different representational systems, any more than the concepts *dog* and *cat*, though undoubtedly different concepts, require radically different representational capacities.

It is often claimed in support of the special nature of *believe*, and in contradiction of the second set of positions above, that solving false belief tasks, such as that in Figure 1, requires that the child to employ a radically different conceptualisation of mental states from that required by understanding *pretend* (e.g., Perner, 1991). Specifically, it is claimed that false belief can only be conceptualised within a "representational theory of mind" (RTM).

I have criticised the RTM view of the preschoolers' theory of mind at length elsewhere (e.g. Leslie & Thaiss, 1992; see also Leslie & Roth, 1993, Leslie & German, in press). Briefly, there are two ways in which one could speak of the child possessing a "representational theory of mind." One could use the term "representational" loosely to cover anything which might be considered true or false, that is, anything which can be semantically evaluated will count as a "representation." In this loose sense of representational theory of mind, mental states involve representations simply because their contents are semantically evaluable and it is irrelevant whether a mental state involves a relation to a proposition or a relation to an image, a sentences, or whatever.

The second and stricter way of using the term is to denote only entities that can be semantically evaluated *and* have a physical form or a syntax. In so far as cognitive science holds a RTM, it is in this second stricter sense of "representational." Thus, for example, psychologists might argue about whether a given piece of knowledge is represented in an image form or in a sentential form. The form of the representation is held to be critical to the individuation of the mental state. Indeed, mental states are individuated within this framework in terms of their form, not in terms of their content. From the point of view of psychology, an image of a cat on a mat counts as a *different* mental state from a sentential representation of the same cat on the same mat.

Because it would be massively confusing to use the same term both for a theory of mind which individuates mental states in terms of their contents (semantics) *and* for a theory of mind which individuates mental states in terms of their form (syntax), we shall use different terms. We shall speak of a "propositional attitude" (PA), or semantically based, theory of mind for the first type of theory and a representational, or syntactically based, theory of mind (RTM) for the second. Now we can ask, Does the child employ a PA based (semantic) theory of mind or representational (syntactic) theory of mind?

Perner's (1991) claim is that success on a variety of false belief tasks at age four reflects a radical theory shift from a PA based theory of mind to a RTM. However, all of the evidence quoted in support of this claim (mainly passing various false belief tasks) only shows that the child individuates beliefs on semantic grounds. After all, the falseness of a belief is a *quintessentially* semantic property. To date, there are no demonstrations of preschoolers individuating beliefs on syntactic grounds in disregard of their content. All the available evidence supports the idea that preschoolers at any rate are developing a semantic theory of belief and other attitudes. What the theory of **ToMM** aims to account for is the specific basis for this early emerging semantic theory of the attitudes.

A task analytic approach to belief problems

How can we begin to investigate the claim that a prestructured representational system interacts with performance factors in producing the patterns seen in the preschool period? Specifically, how do we investigate the notion that performance limitations mask the preschooler's competence with the concept of *belief?* We can try to develop a *task analysis*. In carrying out this analysis, we must separate the various component demands made on conceptual organisation from those made on general problem solving resources in the course of tackling false belief problems.

An important beginning has been made in this line of research by Zaitchik (1990). Zaitchik designed a version of the standard false belief (FB) task in which the protagonist Sally is replaced by a machine, namely, a polaroid camera. The protagonist's seeing the original situation of the marble is replaced by the camera's taking a photograph of it; after moving the marble to a new location, the protagonist's out-of-date belief is replaced in the new task by the camera's out-of-date photograph. While the conceptual content of the task changes (from belief to photograph), the general task

structure remains identical. This task then provides an intriguing control for the general problem solving demands of the FB task. Results from comparing these two tasks show two things: first, normal three year olds fail *both* the FB and the photographs tasks (Zaitchik, 1990), as would be expected on the basis of a general performance limitation; second, autistic children fail only the FB task but pass the photographs version (Leekam & Perner, 1991; Leslie & Thaiss, 1992), consistent with autistic impairment in the domain specific mechanism **ToMM**.

The **ToMM** model can be extended to relate it to the performance limitations affecting young preschoolers. Building on ideas proposed by Roth (Leslie & Roth, 1993; Roth, 1993; Roth & Leslie, in preparation), Leslie & Thaiss (1992) outlined the **ToMM-SP** model. Some false belief tasks, such as the Sally and Ann scenario in Figure 1 and other standards such as "Smarties", make demands on at least two distinct mechanisms. Specific conceptual demands are made of **ToMM** to compute a belief metarepresentation, while, in the course of accurately calculating the content of the belief, more general problem solving demands are made of a "Selection Processor" (**SP**). These latter demands require the child to interrogate memory for the specific information that is key to the belief content inference, disregarding other competing or confusing information. For example, to infer the correct content for Sally's belief, the situation that Sally was exposed to at t_0 has to be identified from memory and the inference to Sally's belief based on that, resisting the pre-potent tendency to simply base the inference upon the (present) situation at t_1.

There is a conceptual basis for the existence of this pre-potent response. *Normatively* beliefs are true: this is what beliefs are "for";—they are "for" accurately describing the world; a belief is "useful" to an Agent only to the extent it is true; in short, beliefs "ought" to be true.[4] It makes sense, then, if, by default, inferences to belief contents are based upon current actuality. In the case of false belief, this normative design fails and in order to accurately compute the errant content, the pre-potent assumption must be resisted. Similar considerations will apply, e.g., to the case of Zaitchik photographs. (The same problem does not arise in the case of true pretends because the Agent is always able to intentionally communicate the content of his pretend whereas, for obvious reasons, an Agent is not in a position to intentionally communicate that he has a false belief. However, see Roth & Leslie [1991] for a case in which communication does help the three year old with false belief.)

We can organise our thinking about the general demands made by some belief tasks by positing a general, or at least non-theory-of-mind-specific, processing mechanism. Roth and I have dubbed this the "Selection Processor" (**SP**). The Selection

---

[4] Although this normative assumption is fundamental to the notion of belief, it is not part of the logic of the concept that belief contents must be true (compare the earlier parallel discussion on page ? of pretends being normatively false). Notice that the normative assumption is a far cry from the "copy theory" of belief (Wellman, 1990). Incidently, it has been my experience that there are many adults who are surprised, even dismayed, to discover that pretends can be true. In view of this, we should not be too hard on the preschooler if she takes a few months to discover that, contrary to design, the vicissitudes of the real world sometimes defeat beliefs with dire consequences for the Agent's goals.

Processor (**SP**) performs a species of "executive" function, inhibiting a pre-potent inferential response and selecting the relevant substitute premise. Like many other "executive functions", **SP** shows a gradual increase in functionality during the preschool period. Some belief tasks do not require this general component or stress it less, by, for example, drawing attention to the relevant "selection" and/or by encouraging inhibition of the pre-potent response. In these cases, better performance on false belief tasks is seen in three year olds (e.g., Mitchell & Lacohée, 1991; Wellman & Bartsch, 1988; Roth & Leslie, 1991; Zaitchik, 1991). According to this view, the three year olds' difficulty with false belief is due to limitations in this general component. Meanwhile, in the normal three year old, **ToMM** is intact.

The autistic child, by contrast, shows poor performance on a wider range of belief reasoning tasks, even compared with Down's syndrome children and other handicapped groups (e.g. Baron-Cohen, 1991c; Baron-Cohen, Leslie, & Frith, 1985; Leslie & Frith, 1988; Roth & Leslie, 1991). This disability is all the more striking alongside the excellent performance autistic children show on out-of-date photographs, maps and drawings tasks (Charman & Baron-Cohen, 1992; Leekam & Perner, 1991; Leslie & Thaiss, 1992). These tasks control for the general problem solving demands of standard false belief tasks. Autistic impairment on false belief tasks cannot, therefore, be due to an inability to meet the general problem solving demands of such tasks. So, although autistic children seem to be impaired in certain kinds of "executive functioning" (Ozonoff, Pennington & Rogers, 1991), this cannot be the cause of their failure on false belief tasks. This pattern can be succinctly explained on the assumption of a relatively intact **SP** together with an impaired **ToMM**—the mirror-image of the normal 3-year-old. Figure 3 summarises the **ToMM-SP** model of normal and abnormal development.

Roth and I have recently extended our approach of studying minimally different task structures in an effort to isolate general processing demands from specific conceptual demands (Roth & Leslie, in preparation). In one study, we compared the performance of young, middle and older three-year-olds on a standard version of the Sally and Ann task with a "partial true belief" version (see Leslie & Frith, 1988 for details of the tasks used). This allowed us to assess the importance of the *falseness* of the belief (a conceptual factor) in generating difficulty for three-year-olds while holding general task structure constant. The results are shown in Figure 4. It can be readily seen that there was no difference in difficulty between the two tasks for three year olds *when task structure is equated.* Taken together with previous findings that three-year-olds can understand "knowing and not knowing" (e.g., Pratt & Bryant,1990; Wellman & Bartsch, 1988), this shows that the conceptual factor of the falseness of the belief *per se* is not the source of difficulty for three year olds. There must be something about the problem solving structure of standard belief tasks that stresses three-year-olds.

Another approach in the literature to the problem of isolating belief competence is to find simplified task structures that 3 year olds perform better on. The theoretical assumption behind such work is that by finding simplified tasks one reduces the number of false negatives that standard tasks produce. Surian and Leslie (in preparation) point out a danger with this approach. Manipulations designed to simplify tasks may inadvertently allow children to pass for the wrong reasons—for reasons which do not

reflect the conceptual competence that the investigator is targeting. To avoid this, we need to introduce controls for false positives. A concrete example will help make the idea of controlling false positives clear. Suppose we run a group of children on the Sally and Ann task and find that 100% of the children pass. We then run another group of same age children on a modified version of the task in which Sally does not go away for her walk but instead remains behind and watches Ann all the while. In this version, Sally sees the transfer of the marble and knows that it is in its new location. Imagine that, to our surprise, we find that 100% in this SEE control group fail! In this control condition, failing means that the child indicates the *empty* location when asked where Sally will look for the marble. Now we will say that the first finding, in the "NOT-SEE" test, of 100% indicating the empty location, consisted of false positives; it did not demonstrate false belief understanding.

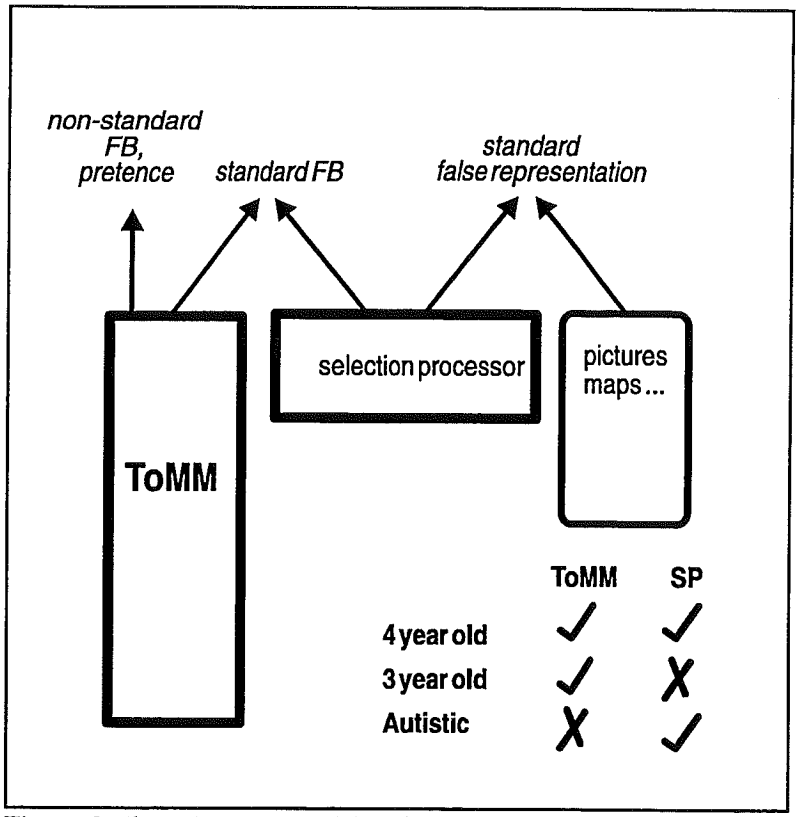

Figure 3. The **ToMM-SP** model of development (after Leslie & Thaiss, 1992).

Suppose instead we had discovered that the false belief group were only 50% correct. We would have been tempted to describe this result as "chance" but we waited till we saw how many children passed the SEE control version of the task. Suppose that on this, 100% of the children succeed. Now we can be confident that in the NOT-SEE test the children really did take Sally's exposure history into account because if they had responded like the controls no-one would have passed: therefore the 50% who did were not false positives. In other words, the second pattern of results says more about false belief understanding than the 100% "passing" in the paragraph above.

One immediate use of this enhanced technique of balancing a NOT-SEE test with a SEE control is to allow us to look at three-year-old performance with a more sensitive instrument.

Robinson & Mitchell (1992) report a study that would benefit from the use of a SEE control. Three year olds were given a scenario in which Sally has two bags each

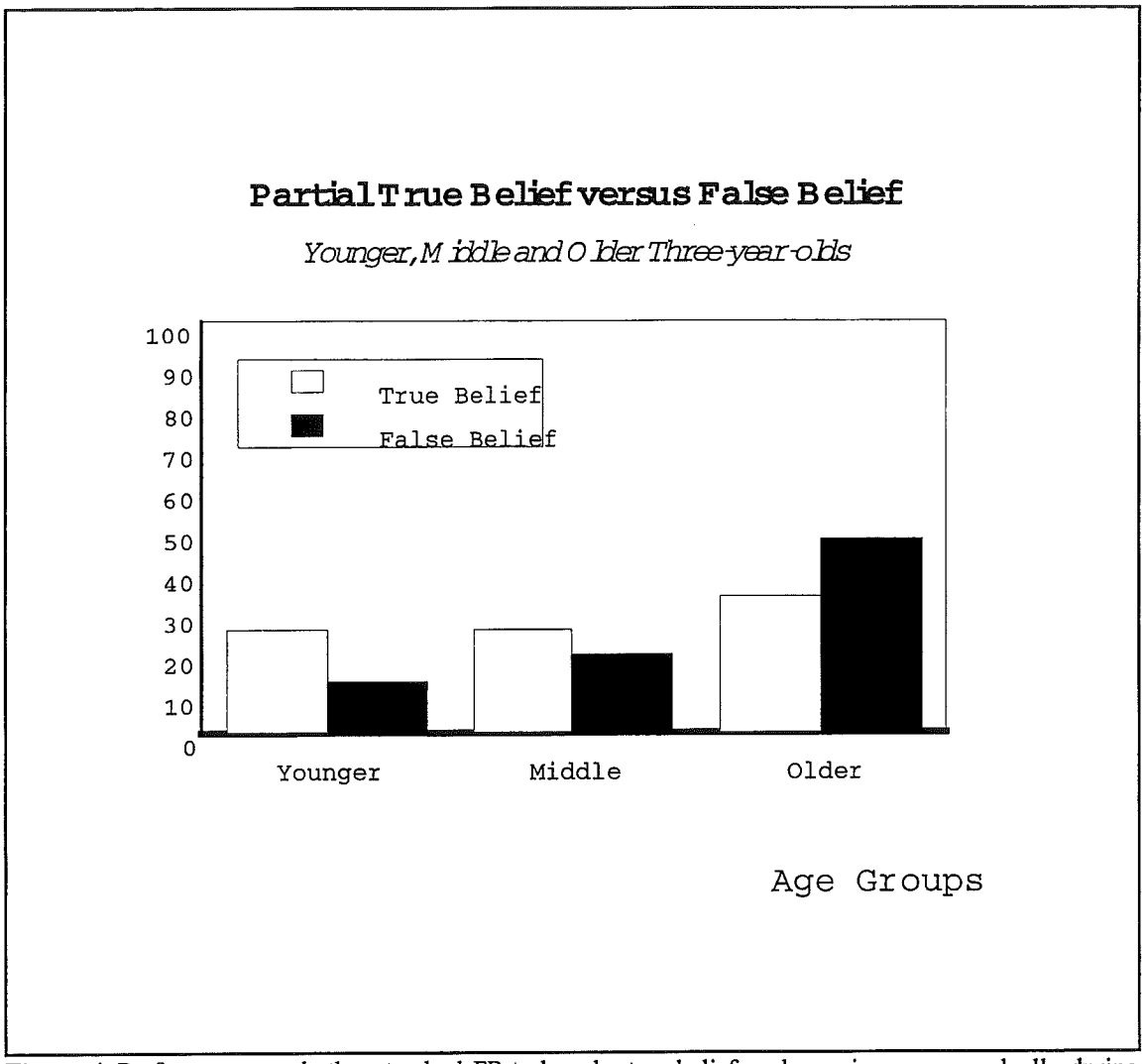

Figure 4. Performance on both a standard FB task and a true belief analogue improves gradually during the fourth year.

containing pieces of material. She places the bags, one in each of two drawers, and goes to the next room. Ann then enters, takes the two bags out of their drawers, plays with them, then replaces the bags. But, by accident, she *swaps* their locations. Sally then calls from the next room that she wants her bag of material to do some sowing and that it is important that she gets the correct bag. She tells Ann that it is the bag in the red drawer she wants but, of course, Sally does not know that Ann has swapped the bags. The child is asked to identify the bag that Sally *really* wants. In this interesting task, the child can correctly identify Sally's desire only if she first relates it to Sally's false belief. The results showed that 50% of the three year olds passed, a higher proportion than that obtained with a standard FB task.

Unfortunately, the possibility exists with this design that children were simply confused by the swapping and interpreted Sally's description of the bag either as

referring to the bag that *was* in the red drawer or to the bag that *is* in the red drawer, half making one interpretation and half the other. This would yield 50% false positives without any of the children actually calculating Sally's false belief. If the children were using such a low level, "dumb" strategy, it will show up again in a version of the Robinson & Mitchell task that implements the SEE control. This time, Sally *watches* as Ann swaps the bags between the drawers. Now, when Sally asks for the bag in the red drawer, she must mean the bag that *is* in the red drawer. Yet, if the children were simply following a "dumb" strategy and not calculating at all what Sally believed, then it should make little or no difference that Sally had watched the proceedings. Half will still interpret her as wanting the bag that was in the red drawer: the confusion created by swapping will occur again, resulting again in 50% "correct". (Bear in mind that the indicated location counted as correct in the SEE control condition is the opposite of that counted correct in the NOT-SEE [false belief] condition).

Surian & Leslie (in preparation) examined the Robinson & Mitchell scenario in relation to the SEE control. They found that the proportion of "correct" locations indicated in the FB condition was the same as the proportion of incorrect locations indicated in the SEEing control. This shows that the children were not taking into account Sally's exposure and thus were providing false positives in the false belief task. In fact, the pattern of responding was very similar to that found on a standard FB task with SEE control. According to these new results, swapping locations and asking about desire in relation to false belief does not, in the Robinson & Mitchell task, produce a simplified false belief problem for three-year-olds.

Surian & Leslie then went on to test three year olds in a study which combines the methods of minimally different task structures with the SEE control for false positives. The Robinson & Mitchell task was modified to introduce *ambiguity* into Sally's desire. Although this modification makes the scenario more complex as a story, we supposed that it would simplify the scenario as a false belief problem. Fodor (1992) has also recently proposed a model of a performance limitation in three-year-old's theory of mind reasoning. In this model, the three-year-old typically predicts behaviour from desire without calculating belief. According to the model, the young child will not calculate belief unless the prediction from desire yields an ambiguous result and the child is unable to specify a unique behaviour. Whenever desire prediction results in ambiguity, however, the child will break the impasse by calculating belief. Standard FB tasks allow unique predictions from desire, so the young child does not calculate belief and thus fails. Older children, however, routinely calculate both belief and desire because they have greater processing capacity available. Like the **ToMM** model, Fodor assumes that the three-year-old possesses the conceptual competence for understanding false belief. Therefore, Fodor predicts that when the three-year-old does calculate belief, she will succeed. One difficulty is to know what predictions the child will consider as ambiguous. For example, though Fodor (1992) suggests splitting and moving the object into two target locations as a way of creating ambiguity, it could be that the child will regard search in *both* locations as a single unambiguous action.

In order to test Fodor's suggestion clearly, we need a scenario in which ambiguity in the object of desire is unavoidable. A modification of the Robinson &

Mitchell scenario meets this requirement nicely. Instead of having two bags of material, Sally has four pencils. Three of the pencils are sharp, while the fourth pencil is broken. Sally leaves the pencils and goes into the next room. Now Ann comes in and finds the pencils. First Ann sharpens the broken pencil, then she breaks each of the original three sharp pencils. Now there are three broken and one sharp pencil. At this point, Sally calls from next door, "Ann, bring me my favourite pencil—you know, it's broken!". As before, the child is asked which pencil Sally really means. The child has been given no information prior to this about which pencil is Sally's favourite. The only information the child has to go on is Sally's attached description, "it's broken". But now there are three pencils which are broken. This unavoidably produces ambiguity. According to Fodor's model, the ambiguity in the object of desire should trigger the child into consulting Sally's belief. When the child consults Sally's *belief* about which pencil is broken, he will realise that Sally thinks that the now sharp pencil is still broken. The *sharp* pencil is the one Sally really means!

Our results showed that 48% of our three year olds correctly chose the pencil that Sally really meant. This performance was significantly better than performance on the unmodified Robinson & Mitchell scenario and better than on the standard false belief task. We thus obtained support for Fodor's ambiguity factor. Before reaching this conclusion, however, we should recognise that there are low level "dumb" strategies that could have produced these results. Perhaps the passers were false positives. For example, the word "favourite" singles out a particular individual. The children may simply have latched onto the "odd-one-out", the uniquely sharp pencil. To control for dumb possibilities like this, we also ran a SEE control version of the pencils task. If children simply respond with this or some other dumb strategy, then they should use the same dumb strategy when Sally remains in the room watching Ann process the pencils. In the SEE control, as in the NO-SEE test, the child has no information on which pencil is Sally's favourite other than the description Sally gives of it as being broken. Again this is ambiguous, because, by the time Sally makes her request, there are three broken pencils. If the children follow a dumb strategy, about half the children should again respond by picking the odd-one-out—the uniquely sharp pencil. In fact, in the SEE control condition only about 15% of the children chose the sharp pencil, the rest choosing one of the broken pencils. This pattern was significantly different from that obtained in the NO-SEE test. Most of the passers were true positives.

By combining a method of minimal task differences with the SEE control, Surian and Leslie obtained a sensitive measure of three-year-old competence. We were able to isolate Fodor's ambiguity of desire factor by comparing the performance on Robinson and Mitchell's original task with the ambiguity modified version of it, while at the same time controlling for false positives by means of a SEE control. Although further studies underway may change the picture, it seems that ambiguity of desire can help three-year-olds in solving a false belief problem.

However, it is not clear that Fodor's model identifies all the performance factors limiting three-year-old's successes. Fodor's (1992) model focused on the prediction of behaviour. Important though this is, the child is also concerned with the underlying mental states themselves. For example, in the ambiguity study above, the child did not

calculate belief in order to predict behaviour. She calculated belief in order to figure out the referent of Sally's desire. Furthermore, in standard FB tasks, even when three-year-olds presumably *do* consult belief—for example, when they are directly requested to do so—they still have difficulty calculating its content accurately. For example, in the standard Sally and Ann scenario it makes little difference if, rather than being asked to predict behaviour, three-year-olds are asked where Sally thinks the marble is. Despite being asked to consult belief, they are no better at calculating its content than when asked to predict behaviour. And even when the ambiguity factor was apparently activated in the study above, still half the children did not calculate belief content correctly.

Fodor's model assumes that three-year-olds do not ordinarily consult belief, but that when they do, they can easily calculate its content correctly (even in the case of false beliefs). The **SP** model, on the other hand, assumes that **ToMM**'s routine calculation of belief normally assumes that content reflects relevant current facts. In light of the normative nature of the belief concept, this assumption is, in the general case, justified. But for false belief situations where belief does not operate as it ought to, in order to produce a correct answer about content, this assumption has to be inhibited or blocked and a specific alternative content identified. Both of these processes (the inhibition of the pre-potent response and the selection of the correct content) stretch the three-year-old's capabilities. Unless "help" is given by the form of the task, three-year-olds will tend to assume beliefs reflect current facts or will fail to identify the correct content.

Fodor's ambiguity of desire can be assimilated to the **SP** model as one factor which can inhibit the normal content assumption and lead to search for an alternative belief content. For example, when the child tries to infer which pencil Sally wants, the first hypothesis will simply be "a broken one". But which one is that, given there are three broken ones? Since it is not possible to reach a definite answer to the question of what Sally wants, the ordinary assumption about belief content is inhibited and an attempt made to calculate the content from Sally's exposure history. Recall that the control children simply had to live with an indefinite answer because in their case Sally in fact knew there were three broken pencils. If the experimental effect simply reflected the dumb strategy, "my first answer is going to be wrong, so I'll pick something else", and the only other different thing the child can pick is the sharp pencil, then this same dumb strategy should have been followed by the SEE control children as well. But it was not. The children were indeed calculating belief. Nevertheless, though the pencils story helped the children, it was not enough to help more than half of them to get their calculation right. Perhaps if task structure were made to help with the selection of the appropriate content as well as with inhibiting the pre-potent response, performance would improve further.

In a final experiment, Surian & Leslie (in preparation) re-examined a modified standard scenario based on Siegal & Beattie (1991). In this otherwise standard Sally and Ann task, instead of asking the child "Where will Sally look for her marble?", the child is asked "Where will Sally look for her marble *first*?". Siegal & Beattie found that adding the word "first" dramatically improved three-year-old's success. This result has

largely been ignored, however, because it is open to some obvious objections. For example, the word "first" may simply lead the child to point to the first location the marble occupied, in other words to follow a dumb associative strategy. Alternatively, the word "first" might cue the child that the experimenter expects the first look to fail. If there is to be a first look, presumably there is to be a second look; but why should there be a second look unless the first one fails? Therefore, point to the empty location for the first failing look. Again, put like this, the word "first" simply triggers a dumb strategy in the child who then appears to succeed but who, in fact, does nothing to calculate belief. We simply added the necessary SEE control to examine the viability of such "dumb strategy" explanations. If the child is not attending at all to Sally's belief then it should make no difference that Sally watches Ann move the marble. In the control condition too, the word "first" should trigger the dumb response strategy.

We found that 83% of the children in the false belief task passed, replicating Siegal & Beattie's finding. If this was the result of a dumb strategy, then we should expect to find a similar proportion *failing* the SEE control task, because in this condition a point to the first location is considered *wrong*. In fact, 88% were correct in this condition too. Therefore, the effect of the word "first" is specific to the status of Sally's belief.

These last results vindicate Siegal & Beattie (1991) and suggest that they have been wrongly ignored. Siegal and Beattie argued that including the word "first" made the experimenter's intentions explicit for the three-year-old. We are now in a position to suggest an account of *how* this manipulation makes the experimenter's intentions explicit for three-year-olds and why they, but not four-year-olds, need the help. Notice that the absolute level of success in the look first task is very high indeed. It is quite comparable to the level of success obtained by four-year-olds in the standard task and higher than that obtained with three-year-olds in the ambiguity task. The word "first" may give the child a "double" help. The child's attention is drawn to the possibility that Sally's first look may fail. If her first look fails (to find the marble), Sally's behaviour defeats her desire. Plausibly, behaviour defeating a desire encourages the blocking of the normal assumption regarding belief content in the same way that ambiguity about what would satisfy the desire does. This is a variation on Fodor's factor. In addition, however, the word "first" directs the child's attention to the first location and this helps select the correct counterfactual content. The double help results in very good performance. Finally, bear in mind, once again, that this hypothesised double help has specific effects depending upon the status of Sally's belief. If Sally indeed knows where the marble is, asking where she will look *first* obtains the contrasting answer from three-year-olds: namely, "in the second location".

In summary, there is increasing evidence that three year olds have an underlying competence with the concept of belief but that this competence is not revealed in the tasks that are standardly used to tap it. It seems increasingly likely that their competence is masked by a number of "general" factors that create a performance squeeze. This squeeze gradually relaxes over the course of the fourth year (see Figure 4), and probably beyond. Further support is provided for this general view by the finding that false belief tasks show a *difficulty gradient*, that some false belief tasks are

easier than others. The **ToMM-SP** model provides, to date, the most wide ranging model of the young child's normal theory of mind competence, of the performance factors that squeeze the child's success on false belief calculations, and of the abnormal development of this domain found in childhood autism.

References

Baldwin, D.A. (in press) Infants' ability to consult the speaker for clues to word reference. *Journal of Child Language.*

Baron-Cohen, S. (1991) The development of a theory of mind in autism: Deviance and delay? In M. Konstantareas and J. Beitchman (Eds.), *Psychiatric Clinics of North America*, Special issue on *Pervasive developmental disorders*, (pp. 33–51). Pennsylvania: Saunders.

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

Charman, T., & Baron-Cohen, S. (1992) Understanding drawings and beliefs: A further test of the metarepresentation theory of autism (Research Note). *Journal of Child Psychology and Psychiatry*, **33**, 1105–1112.

Flavell, J.H. (1988) The development of children's knowledge about the mind: From cognitive connections to mental representations. In J.W. Astington, P.L. Harris, & D.R. Olson, (Eds.), *Developing theories of mind*, (pp. 244–267). New York, NY: Cambridge University Press.

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

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

Gopnik, A. & Slaughter, V. (1991) Young children's understanding of changes in their mental states. *Child Development*, **62**, 98–110.

Grice, H.P. (1957) Meaning. *Philosophical Review*, **66**, 377–388.

Harris, P.L., & Kavanaugh, R. (in press) The comprehension of pretense by young children. *Society for Research in Child Development Monographs.*

Leekam, S., & Perner, J. (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". *Psychological Review*, **94**, 412–426.

Leslie, A.M. (1988a) Causal inferences in shared pretence. Paper presented to BPS Developmental Conference, Coleg Harlech, Wales, 16-19th September 1988.

Leslie, A.M. (1988b) Some implications of pretense for mechanisms underlying the child's theory of mind. In J. Astington, P. Harris & D. Olson (Eds.), *Developing theories of mind*, (pp. 19–46). Cambridge: Cambridge University Press.

Leslie, A.M. (1988c) The necessity of illusion: Perception and thought in infancy. In L. Weiskrantz (Ed.), *Thought without language*, (pp. 185–210). Oxford: Oxford Science Publications.

Leslie, A.M., & Frith, U. (1988) Autistic children's understanding of seeing, knowing and believing. *British Journal of Developmental Psychology*, **6**, 315–324.

Leslie, A.M., & Frith, U. (1990) Prospects for a cognitive neuropsychology of autism: Hobson's choice. *Psychological Review*, **97**, 122–131.

Leslie, A.M., & German, T. (in press) Knowledge and ability in "theory of mind": One-eyed overview of a debate. In M. Davies (Ed.), *Folk psychology: Simulation or theory?* Oxford: Blackwell.

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.

Marr, D. (1982) *Vision.* San Francisco: W.H.Freeman & Co.

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

Ozonoff, S., Pennington, B.F., & Rogers, S.J. (1991) Executive function deficits in high-functioning autistic individuals: Relationship to theory of mind. *Journal of Child Psychology and Psychiatry*, **32**, 1081–1105.

Perner, J. (1988) Developing semantics for theories of mind: From propositional attitudes to mental representation. In J. Astington, P.L. Harris & D. Olson (Eds.), *Developing theories of mind*, (pp. 141–172). Cambridge: Cambridge University Press.

Perner, J. (1991) *Understanding the representational mind.* Cambridge, MA: MIT Press.

Pratt, C., & Bryant, P. (1990) Young children understand that looking leads to knowing (so long as they are looking into a single barrel). *Child Development*, **61**, 973–982.

Premack, D. & Woodruff, G. (1978) Does the chimpanzee have a theory of mind? *The Behavioural and Brain Sciences*, **4**, 515–526.

Pylyshyn, Z.W. (1978) When is attribution of beliefs justified? *The Behavioural and Brain Sciences*, **1**, 592–593.

Quine, W.V. (1961) *From a logical point of view*. Cambridge, MA: Harvard University Press.

Robinson, E.J., & Mitchell, P. (1992) Children's interpretation of messages from a speaker with a false belief. *Child Development*, **63**, 639-652.

Roth, D. (1993) *Beliefs about false beliefs: Understanding mental states in normal and abnormal development.* Ph.D. Thesis, Tel Aviv University.

Roth, D., & Leslie, A.M. (1991) The recognition of attitude conveyed by utterance: A study of preschool and autistic children. *British Journal of Developmental Psychology*, **9**, 315–330. Reprinted in G.E. Butterworth, P.L. Harris, A.M. Leslie & H.M. Wellman (Eds.), *Perspectives on the child's theory of mind*, (pp 315–330). Oxford: Oxford University Press.

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

Vygotsky, L.S. (1967) Play and its role in the mental development of the child. *Soviet Psychology*, **5**, 6–18.

Wellman, H.M. (1990) *The child's theory of mind.* Cambridge, MA: MIT Press.

Wellman, H.M., & Bartsch, K. (1988) Young children's reasoning about beliefs. *Cognition*, **30**, 239–277.

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.

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

Zaitchik, D. (1991) Is only seeing really believing?: Sources of true belief in the false belief task. *Cognitive Development*, **6**, 91–103.