Segal, G. (1997). *Representing representations*. Unpublished paper, University College, London.
Shatz, M., Wellman, H., & Silber, S. (1983). The acquisition of mental verbs: A systematic investigation of first references to mental state. *Cognition, 14*, 301–321.
Siegal, M., & Beattie, K. (1991). Where to look first for children's knowledge of false belief. *Cognition, 38*, 1–12.
Siegal, M., & Peterson, C.C. (1994). Children's theory of mind and the conversational territory of cognitive development. In C. Lewis & P. Mitchell (Eds.), *Children's early understanding of mind: Origins and development* (pp. 427–455). Hove, UK: Lawrence Erlbaum Associates Ltd.
Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition*. Cambridge, MA: Harvard University Press.
Steeds, L., Rowe, K., & Dowker, A. (1997). Deaf children's understanding of beliefs and desires. *Journal of Deaf Studies and Deaf Education, 2*, 185–195.
Sullivan, K., & Winner, E. (1991). When 3-year olds understand ignorance, false belief and representational change. *British Journal of Developmental Psychology, 9*, 159–171.
Tager-Flusberg, H. (1994). What language reveals about the understanding of minds in children with autism. In S. Baron-Cohen, H. Tager-Flusberg, & D.J. Cohen (Eds.), *Understanding other minds: Perspectives from autism* (pp. 138–157). Oxford: Oxford University Press.
Tager-Flusberg, H. (1995). *Language and the acquisition of a theory of mind: Evidence from autism and Williams syndrome*. Paper presented at the biennial meeting of the Society for Research in Child Development, Indianapolis, IN.
Tager-Flusberg, H. (1996). *Relationships between language and thought: Cognition verbs and theory of mind*. Paper presented at the meeting of the International Association for the Study of Child Language, Istanbul, Turkey.
Tager-Flusberg, H. (1997). The role of theory of mind in language acquisition: Contributions from the study of autism. In L. Adamson & M.A. Romski (Eds.), *Research on communication and language disorders: Contributions to theories of language development*. Baltimore, MD: Paul Brookes Publishing.
Tager-Flusberg, H., & Sullivan, K. (1994). Predicting and explaining behavior: A comparison of autistic, mentally retarded and normal children. *Journal of Child Psychology and Psychiatry, 35*, 1059–1075.
Wellman, H.M. (1990). *The child's theory of mind*. Cambridge, MA: MIT Press.
Wimmer, H., & Hartl, M. (1991). Against the Cartesian view on mind: young children's difficulty with own false beliefs. *British Journal of Developmental Psychology, 9*, 125–138.
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.

CHAPTER ELEVEN

Attending to and learning about mental states

Tim P. German
Department of Psychology University of Essex, UK

Alan M. Leslie
Center for Cognitive Science, Rutgers University, USA

OVERVIEW

A current view in the 'theory of mind' literature is that children acquire knowledge about the mind between the ages of 2 and 5 years by going through a process of conceptual 'discovery'. Most theorists who make this assumption couch this discovery process in terms of the child actively constructing a kind of 'scientific theory' (Gopnik, 1993, 1996a,b; Gopnik & Meltzoff, 1997; Gopnik & Wellman, 1992, 1994; Perner, 1988, 1991, 1995). To date, none of these theorists have provided an account of (i) how even a single mental state concept might be 'discovered' this way, (ii) how any of the proposed sequences of theories might be 'constructed', (iii) what the critical evidence is that children are exposed to at different points in the sequence, or (iv) how children are able to identify and assess this 'evidence' as critically relevant to the process of theory discovery. In general, although a 'theory' is an information structure, theory-theorists seem unaware of the information processing issues raised by their proposals.

We explore an alternative approach that rejects the assumption that concept possession depends upon discovering theories. Instead, we give a central role to an elementary ability to attend to mental states. We examine how the elementary ability to attend to *beliefs* becomes gradually more flexible and promotes increasing success on false belief problems. Using examples from the literature, together with new evidence on children's reasoning about different mental states, we illustrate the larger role that these specialised attentional mechanisms play in allowing learning about mental states to take place.

INTRODUCTION

In making sense of the actions, reactions and interactions of other agents, appeal to abstract concepts such as BELIEF, DESIRE and PRETEND, comes so naturally as to appear trivial.[1] Indeed, this knowledge and ability sometimes goes under the name of mere 'common-sense' or 'intuitive' psychology, and one might be forgiven for thinking therefore, that there must be other, more rigorous, more scientific models of mind or brain that provide better or more successful interpretations, predictions, and explanations of behaviour. But as Pinker (1997, p. 63) reminds us:

> intuitive psychology is still the most useful and complete science of behaviour there is . . . this part of common sense has so much power and precision in predicting, controlling and explaining everyday behaviour, compared to any alternative ever considered.

It is striking to reflect that the best efforts of several generations of the brightest research scientists have done so little to improve upon the ideas that are grasped effortlessly by every untutored 4-year-old, mostly before they can add two and two.

The theorist attempting to explain how this common-sense knowledge is acquired is faced with many puzzles. Leslie (1994a) describes the 'fundamental problem of theory of mind' as follows: given that beliefs, desires and pretends can be neither seen, heard nor felt, how does the young brain succeed in learning about them? In this chapter, we review two principal kinds of answer provoked by this question. One answer suggests that 'theory of mind' knowledge is acquired by discovering a succession of explicit 'theories' about mental states, through the operation of 'theory'-formation processes akin to those involved in 'real' science formation. The second kind of answer suggests that innate mechanisms provide an initial representational competence—a small, core set of primitive mental state concepts that allow the child to attend to instances of a basic set of mental states, and thus *to learn about them*.

The role of 'theory' in 'theory of mind' has been greatly exaggerated, or so we believe. One unfortunate consequence has been to downplay the role of performance factors in 'theory of mind' development. We will argue that it is necessary to understand the role of performance as well as competence factors, and will outline several strands of recent evidence that highlight the need to develop information processing models of how conceptual competence is deployed and how this deployment develops.

The theory-theory account of common-sense psychology

According to the theory-theorists, children 'discover', over the pre-school period, a succession of 'theories' about the mind (Forguson & Gopnik, 1988; Gopnik, 1993, 1996a, b; Gopnik & Meltzoff, 1997; Gopnik & Wellman, 1992, 1994; Perner, 1988, 1991, 1993, 1995; Perner, Baker, & Hutton, 1994; Wellman, 1990). The culmination of this process, at about age four, is the 'adult theory' that reveals the *representational nature of beliefs*. The emergence of the representational 'theory' is marked by the child succeeding at standard false belief tasks (Baron-Cohen, Leslie, & Frith, 1985; Perner, Leekham, & Wimmer, 1987; Wimmer & Perner, 1983). Children begin to predict that a protagonist, who has hidden an object in a particular location then been absent while it was moved elsewhere, will subsequently search for it where she *thinks* it is, rather than where it really is. Younger children typically fail this task by predicting search at the object's current location. The theory-theory explains the shift in performance between age three and four by a change in the child's conceptual competence, which, in turn, is explained by a shift in the underlying 'theory' the child holds with regard to the nature of mental states. The younger child fails because she operates with the wrong 'theory' of belief which dictates that she employ a deficient BELIEF concept. The older child, having grasped the adult 'theory', now employs the adult concept which bestows the ability to solve the false belief task.

'Theory' plays the central role in 'theory-theory' because which concepts you possess depends upon which theories you grasp. The theory-theorist is thus committed to offering a characterisation of the content of each of the theories (said to be) acquired by the child in the course of 'theory of mind' development. The theory-theorist also needs to address the question of how the successive theories are acquired, and thus how children discover concepts such as BELIEF, DESIRE and PRETEND. It will be our contention that, to date, no plausible characterisation of theory contents nor serious account of 'theory' formation has emerged.

Two versions of the theory-theory can be identified. Though they have much in common, and many of the arguments raised here will apply to both, it is worth outlining the principal differences between these versions. Perner (1988, 1991, 1995; Perner et al., 1994) is concerned primarily with accounting for the shift in children's reasoning about the mind between ages 3 and 4. He makes specific proposals about the nature of both the 3-year-old theory, and about the content of 4-year-olds' representational theory of belief. Though Perner (1995, p. 264) sees the shift between the two as involving theory change:

> my view lies in the tradition of the theory-formation view of conceptual change.

Perner (1995, p. 264) distances himself from making any strong claims with respect to how the conceptual changes are realised:

> The use of the word *theory* is to emphasize that concepts develop within groups of theoretically related concepts. Its purpose is not to model children's intellectual growth on the way in which scientists at the forefront of human knowledge break new ground.

For Perner, then, 'theory-theory' is primarily the view that concept possession, and thus concept change, depends upon grasping the 'theory' that defines the concept. He remains sceptical about whether the process through which a child forms a 'theory' is really the same as the process through which a scientist forms a theory.

By contrast, Gopnik and her colleagues (Gopnik, 1993, 1996a, b; Gopnik & Meltzoff, 1997; Gopnik & Wellman, 1992, 1994) have no such qualms (but see Wellman, 1990, p. 130 for a dissenting view). These writers (Gopnik, 1996a, p. 169) propose a deeper explanatory role for the 'child as scientist'[2] than mere analogy:

> there are quite distinctive and special cognitive processes that are responsible both for scientific progress and for particular kinds of development in children.

Gopnik's position also differs from Perner's in terms of the specific nature of the successive theories of mind that children are said to discover. First, children adopt a 'belief as copy' theory at age 3 (Gopnik & Wellman, 1992, 1994; Wellman, 1990), rather than the 'situation theory' (or 'theory of prelief') proposed by Perner (1991, 1995; Perner et al., 1994). Secondly, Gopnik (and Wellman) is far less explicit than Perner about what the content of the 4-year-old's 'representational theory of belief' is supposed to be. In fact, it is hard even to determine whether Gopnik's (and Wellman's) 'representational theory of belief' is the same or a different 'representation theory of belief' that Perner attributes to the 4-year-old. Thirdly, the two versions have somewhat different views on the extent to which later 'theories' of mind replace earlier versions rather than simply extend them (see for example Perner, 1991, chapter 10 versus Gopnik, 1993).

In the following sections, we outline some of the more important details of these two versions of theory-theory. Following this, we discuss some general problems with theory-theories, as well as some more specific problems that arise for each version. Finally, we outline an alternative to theory-theory, along with several strands of evidence in its favour.

Understanding semantic evaluation and the representational theory of mind (RTM)

Perner's focus over the last several years has been on the specific conceptual shift that the child makes between age three and four, that results in the child becoming able to solve the false belief task (Perner, 1988, 1991, 1993, 1995; Perner et al., 1987; Perner et al., 1994; Wimmer & Perner, 1983). Perner has described the key conceptual achievement that enables an understanding of false beliefs as the child grasping the defining feature of representation. That is, as Perner (1995, p. 248) stated, the child 'must have mastered the distinction between what something represents (refers to, makes a claim about) and how it represents (conceives of) it as being'. In the case of the false belief task, this distinction is between the target of Sally's belief (the current location of the marble) and the way that Sally represents this target as being (a non-factual 'situation' where the marble remains in the basket).[3]

Children must have, at minimum, mastered the above distinction in a very particular way to have what Perner (1995, p. 248) describes as 'a representational understanding of mind'. For Perner, it is *not* enough for the child merely to have an *ability* to honour this distinction as she represents people's mental states. An ability merely to honour this distinction while representing pretence was built into the metarepresentational mechanisms proposed by Leslie (1987) and roundly criticised by Perner (1991) as entirely inadequate for genuine metarepresentation, for the sort of metarepresentation required for understanding belief. For Perner's metarepresentation, the child must have explicit knowledge that there is such a distinction and that it is precisely this distinction that the concept BELIEF expresses. We can make the difference between Leslie and Perner a little clearer by considering an example that does not have the confusing feature that we are discussing a mental entity (BELIEF) that designates a mental entity (belief). We can all agree that having the concept WATER is having a concept that expresses the property of *being* H_2O. The equivalent of Leslie's position is that you have the concept WATER just in so far as your concept WATER in practice represents things that have the property of *being* H_2O. You are not required to know what H_2O is nor that H_2O is the property that WATER expresses. That is, you are not required to know the chemical theory of water (or any other theory) in order to have the concept WATER. The equivalent of Perner's position, is that you cannot have the concept WATER unless and until you understand that water is really H_2O—coming to possess the concept WATER is discovering the theory that water is really H_2O. Our hunch is that most human beings, past and present, possess the concepts WATER and BELIEF without understanding either chemical theory or the representational theory of mind, respectively.

As well as explicit claims about the 4-year-old's theory of mind, Perner has specified in some detail the 'theory' held by the younger child. He has described this variously as a 'propositional attitude' theory (Perner, 1988), a 'situation' theory (Perner, 1991, 1993) and most recently a 'prelief' theory (Perner, 1995; Perner et al., 1994). In a recent defence of his position (Perner, 1995), younger children are characterised as able to relate an agent (in a particular mental state) to a 'situation'.[4] The child herself might evaluate the 'situation' as true or false, in Perner's notation, child represents: [Agent—attitude → 'proposition'], and child represents the proposition (P) as being true or false. The additional insight gained by 4-year-olds is to recognise that situations are *also evaluated by the agents holding attitudes to them* and *might be evaluated differently by different agents*. Perner captures this in his notation as: [Agent—attitude→(P), Agent—semantically evaluates as true→(P), P is false].

The understanding attributed to the 3-year-old is sufficient, according to Perner, to explain young children's success at reasoning about concepts such as DESIRE (Hadwin & Perner, 1991; Wellman & Woolley, 1990). A desire is represented as the agent desiring a state of affairs: [Agent—wants → (P)], and if the child evaluates the state of affairs as true then the desire is understood to be fulfilled, while if the state of affairs is evaluated as false the desire is understood as unfulfilled.

Perner also claims that 3-year-old children, with this conceptual repertoire, can understand something (though not all) of the 'belief component' of adult belief-desire psychology. This is evidenced by their appreciation of pretend play, understood by the younger child as an agent acting according to propositions that the child evaluates as false. The child can achieve this understanding, according to Perner, because she is able to relate the agent to the proposition and evaluate the proposition as false [Agent—attitude → (P), P is false]. But, because she cannot understand that the agent also semantically evaluates the proposition (and may evaluate it differently from the child), she cannot distinguish this pretend action from an inappropriate action caused by a false belief. Accordingly, Perner suggests that the younger child's understanding of the epistemic component of belief-desire psychology takes the form of an undifferentiated understanding of belief and pretend, so called PRELIEF, and hence Perner's use of the phrase 'prelief theorist' alongside 'situation theorist', to describe children at this stage. Elsewhere (Perner et al., 1994: footnote p. 284), he states that understanding the mind via the concept PRELIEF is to be read as compatible with his ideas that younger children are *situation theorists* (Perner, 1991). Perner has also talked of the younger child's competence as involving 'understanding propositional attitudes' (Perner, 1988, though see endnote no. 4, p. 169).

Perner seems torn between two contradictory views of early pretence understanding. He is tempted in places (Perner, 1988, 1991) to attribute to the young child a purely behaviourist 'theory' of pretence as merely a form of action. At other times, he is willing to attribute an early understanding of pretence (or 'prelief') as a propositional attitude in which an agent pretends a fictitious something (e.g. "that's a telephone") about a real something else (e.g. about a real banana). But here the child's processing 'merely honours' the fact that the agent pretends this of that and does not represent explicitly what pretending (or 'prelieving') really is, namely a relation in which an agent semantically evaluates a mental representation that expresses a proposition.

We can easily agree with the second of the above views of early pretence understanding. Our only difference on this score is that we think (a) that nothing more than 'honouring' is required in order to understand belief, and (b) that young children have two distinct concepts, PRETEND and BELIEVE, instead of a single undifferentiated concept, PRELIEF.

The 'child as scientist'?

The key claim made by proponents of the 'child as scientist' version of the theory-theory, is that the very same theory-formation abilities responsible for our cultural scientific progress are at the heart of children's acquisition of knowledge about the mind and knowledge in other domains (Gopnik, 1993, 1996a, b; Gopnik & Meltzoff, 1997; Gopnik & Wellman, 1992, 1994). As Gopnik (1996a, p. 169) puts it:

> theories and theory changes, in particular, are responsible for the changes in children's understanding of mind.

According to this version of theory-theory, children are equipped with innate 'starting-state' theories which are modified and revised by innate theory formation mechanisms, with such modification argued to start as early as birth (Gopnik 1996a, b; Gopnik & Meltzoff, 1997). Children's performance on various mental state reasoning problems, on this view, is to be accounted for in terms of the properties of successive theories. Again, a key assumption is that the concepts you have depend upon the theories you grasp. The child employs different concepts at different stages in development because new theories replace older theories. There are even 'transitional' periods in the individual's life that are akin to periods in the history of our culture during which a scientific 'paradigm' shifts. Gopnik and Wellman (1992, p. 156) put it thus:

> Recent evidence suggests that during the period from three to four many children are in a state of transition between the two theories, similar, say to the fifty years between the publication of *De Revolutionibus* and Kepler's discovery of elliptical orbits.

Gopnik and Wellman (1992, 1994) describe a succession of theories of the mind employed by children between age 2 and age 5. The 2-year-old theory recognises fewer, simpler, different concepts than do later theories; DESIRE and PERCEPTION are 'drives toward' objects and 'awarenesses of' objects respectively: 'in neither case need the child conceive of a complex propositional relationship between these mental states and the world' (Gopnik & Wellman, 1992, p. 150). The 2-year-old theory is revised such that a successor theory (e.g. at age three) recognises new concepts, for example a version of BELIEF modelled on earlier understanding of perception: 'belief, like perception and desire, involves rather direct causal links between object and believers' (ibid., p. 151). There is still more conceptual change required however, because this early 3-year-old concept of BELIEF does not allow for false contents.

It is only after children acquire a 'quite different view of mind' (ibid., p. 152), at around age 4 or 5, that their concepts resemble the adult versions: 'perceiving

becomes perceiving that and desiring becomes desiring that, we might even add, that believing becomes believing that' (ibid., p. 153). The key achievement, as we saw with Perner's position, is the mastery of the concept of *mis*representation allowing success at the false belief task. Only then is the child credited with having constructed a 'representational theory of mind'.

Contrary to the claims of Gopnik (1993), the 'copy theory' of belief cannot be the same as the 'situation theory' or the 'prelief' concept proposed by Perner (1988, 1991, 1995; see above). The child who understands beliefs only as copies of the world has a different understanding than a child who can conceive of agents as holding attitudes to 'situations' that might be evaluated by the child as either true or false (Perner, 1995). If a child were to understand beliefs as *only being true*, then, together with Perner (1991), one must ask, why is this not simply the concept, KNOWING?

As Fodor (1992) points out, Wellman's own data appear to show that 3-year-old children are not limited to a 'copy theory of belief', i.e. to the concept, KNOW; for example, in his 'Not Own' belief task (Wellman & Bartsch, 1988), young children understand that Billy might think his puppy is in the garage and will look for it there when the child herself guesses it is in the shed, and knows that she would look in the shed. These beliefs cannot *both* be copies of reality (see also Perner, 1991, pp. 275–276).

Wellman also has data that contradict the 'drive theory of desire' that he attributes to 2-year-old children. Children in Wellman and Woolley (1990) were required to understand that Billy, for example, wants to find a puppy so that he can take it to school. They found that the children were capable of predicting that Billy will cease searching and go to school with his puppy when he finds his puppy, but not when he finds nothing and not when he finds a different pet. Wellman's theory characterises the 2-year-old's 'theory' of desire as a drive toward a physical object. Wellman proposes this 'drive theory' to reconcile the data showing very early understanding of desire with his larger assumption that children younger than 4 years cannot understand propositional attitudes. Wellman's move is to attribute to the young child only mental state notions that relate agents to physical objects and not to states of affairs. This is because he assumes that a state of affairs would have to be represented by the child as a proposition (see for example, Wellman & Woolley, 1990, p. 249). Thus, to attribute mental state notions that relate agents to states of affairs would be, Wellman thinks, to attribute propositional attitudes to young children, something he rules out on a priori grounds. He hopes to avoid this result by attributing to young children mental state notions in which agents relate only to physical objects, presumably because physical objects can be represented by single concepts rather than by propositions.

But the 2-year-olds in Wellman's experiment correctly predict that Billy will go to school when he finds his puppy, a prediction based on understanding Billy's desire to find his puppy *in order that he can take it to school*. The children have understood that Billy's finding the 'object' satisfies a sub-goal of a larger goal. But this larger goal is expressible only as a state of affairs. Therefore, the child cannot be limited to conceiving of desire as 'an internal longing for an external object', if 'object' means physical object as opposed to state of affairs, because 'he will take his pet to school' is not a physical object, but a state of affairs. Wellman's own data therefore contradict his 'drive theory of desire' (see also German, 1995a; Leslie, 1994b; Perner, 1991).

Problems with the theory-theory

Both versions of the theory-theory claim that *conceptual change* in the domain of common-sense psychology are responsible for the pattern of performance on theory of mind tasks over the first few years of life. But precisely what kinds of conceptual changes are hypothesised to occur? What specific mechanisms are proposed to realise theory change?

The 'child as scientist' view advocated by Gopnik and colleagues appears at first blush to offer an answer to this question; the mechanisms that come into play are the same mechanisms that are responsible for theory changes in science. But nothing is known about the mechanisms of conceptual change in science. Gopnik's 'mechanisms' are simply a reiteration of familiar, but still mysterious, processes like 'testing predictions', 'evaluating evidence', the 'positing of hypotheses' and so on, the stuff of popular conceptions of science. However, she makes no new proposals about where new concepts come from, how the concepts that feature in hypotheses under test get to be there in the first place, how 'evidence' leads to new 'theories', and so on. Apparently, Gopnik believes that addressing these information processing questions is unnecessary (Gopnik, 1993).

Interestingly, there are other accounts of conceptual change in the literature in regard to domains other than theory of mind. Carey and colleagues (Carey, 1985, 1988, 1991, 1995; Carey & Spelke, 1994, 1996; Johnson & Carey, 1998) have suggested that conceptual change is involved in the restructuring of children's knowledge about living things between the ages of 4 and 10 years, changes in children's intuitive theory of matter between ages 6 and 12 years, intuitive cosmology (Vosniadu & Brewer, 1992), as well as the construction of certain numerical concepts such as zero, infinity, negative numbers and fractions during the later elementary school period (Gelman, 1991). Although Gopnik sees the work of these authors as supporting the general theory-theory line on 'theory of mind' (Gopnik, 1996a, p. 169, 1996b, p. 485; Gopnik & Wellman, 1992, p. 145, 1994, p. 257), there are important differences between the approach taken by Gopnik and these other uses of the theory-theory framework.

First, the principal claims about conceptual change in these other domains are seen as descriptive rather than explanatory (see, for example, Carey, 1988, p. 141); the issue for Carey is the extent to which it might be useful to describe changes in a child's knowledge in a given domain as changes in an intuitive

'theory'. The relationship between the concepts and structure of the 'theory' at one point in time (T_1) and the 'theory' at a later point in time (T_2) can be described as involving conceptual changes of specific types. For example, *differentiation* where a concept recognised in T_1 is split into two concepts in T_2, *coalescence* where more than one T_1 concept is aggregated to just one concept in T_2, concepts can change in their peripheral/core feature structure, such that changes to a concept core are regarded as a conceptual change, and, finally, conceptual change can result in the creation of new ontological categories (Chi, 1992). Perhaps this kind of description can be offered for the putative 'theories' in the domain of theory of mind: BELIEF and PRETEND might be differentiated from PRELIEF (Perner, 1995), or there may be a shift from understanding pretend in terms of peripheral or characteristic features (action) to understanding the core or defining feature of mental representation (Lillard, 1993).

However, where putative mechanisms of change have been discussed (Carey & Spelke, 1994), they have been argued to involve complex processes such as the creation of mappings across domains, physical analogies, thought experiments and limiting case analyses. Johnson and Carey (1998) also discuss, in general terms, the types of abilities upon which conceptual change might be based. They suggest, in addition to the above, processes such as meta-cognitively aware uses of analogy, comprehension monitoring and abstract same/different comparisons. There are still no details of what these theory formation mechanisms might look like, but a key point is that such abilities seem themselves to develop after the pre-school period during early school age, and to be impaired in those with mental retardation (Campione, Brown, & Ferrara, 1982).

By contrast, as we indicated above, proponents of the theory-theory in the domain of theory of mind have not addressed the question of underlying mechanisms. Moreover, it is unlikely that the kinds of processes that have been proposed for theory change in other domains could account for conceptual change in theory of mind. Theory of mind acquisition proceeds uniformly, routinely, effortlessly, in the first few years of life, is very largely independent of IQ (Leslie, 1987), and does not require formal instruction. All of these differences between 'theory of mind' knowledge and common-sense knowledge in other domains (such as intuitive biology, theory of matter and number) have led even proponents of conceptual change in other domains to doubt that the theory-theory approach is appropriate to explain the acquisition of common-sense psychology (see for example Carey & Spelke, 1996; see also German & Samuels, in preparation, for further discussion of these issues).

Turning to Perner's account, recall that there is no commitment to the 'child as scientist' as anything other than analogy. For Perner, the process of acquisition does not have to mirror what a professional scientist does; instead, the role of theory-theory is to give an account for the nature of concepts and concept acquisition. The child comes to possess knowledge of a domain that is theory-like in the following sense: she has to discover the nature of the things in the domain—she has to discover what these things really are. In particular, for the 'theory of mind' domain, the child must discover what beliefs (and other mental states) really are, namely, *representations*. This gives Perner's approach greater explanatory depth than the Gopnik–Wellman version of theory-theory. By insisting that the child must develop a 'theory' of what beliefs really are, he is able to account for why grasping *that* 'theory' results in the possession of a particular concept, namely, BELIEF, and to establish criteria whereby we may judge whether or not the child in fact possesses the concept.

Perner's views are also distinguished from those of Gopnik–Wellman because he has been at pains to specify what the 'theory' is that he is attributing to the child. Clearly, if, as according to theory-theory, the key to understanding development is that the child is supposed to grasp and hold a certain 'theory', then we are entitled to be told exactly what this 'theory' is that the child grasps and holds. Neither Gopnik nor Wellman ever tell us what the child's 'representational theory' of belief is. But Perner does. According to Perner (1995), the 4-year-old child comes to understand that when John believes a certain proposition *p*, John is semantically evaluating a mental representation which expresses *p*. As Roth & Leslie (1998) point out, stating the child's 'theory' inevitably means using other concepts. And because the 'theory' stated is something that the child is said to grasp, this means that the child must possess those concepts too. Therefore, in order to acquire the concept, BELIEF, the 4-year-old, according to Perner, must acquire the prior concepts, SEMANTIC, EVALUATE, MENTAL, REPRESENTATION, EXPRESS and PROPOSITION. The natural next question is, how are *these* concepts acquired?

Perner expresses a leaning toward the view that 'theory of mind' is acquired through 'interaction with the environment' (1995, p. 263). We have no doubt that 'interaction with the environment' plays an important role; in fact, we regard this essentially as a truism. The question is, what role precisely? It is not at all easy to imagine how the six concepts comprising the child's 'representational theory of belief' are acquired through 'interaction with the environment'.

In summary, current theory-theory views suffer from a number of shortcomings which invite a more sceptical attitude than is usually given them in the field. Although future development of theory-theory may overcome some of these problems, our feeling is that the entire approach is flawed. The heart of this flaw, in our view, is the assumption that possession of abstract concepts *must* depend upon possession of abstract knowledge.

INFORMATION PROCESSING AND THE THEORY OF MIND MECHANISM

We turn now to consider an alternative to the conceptual change framework for the acquisition of knowledge in this domain. We believe that understanding children's developing theory of mind means specifying the mechanisms that will

allow the child firstly to *attend to* and subsequently to *learn about* mental states. Without the ability to attend to the relevant properties in the environment, there is no way that the child will learn about states such as *beliefs*, *desires* and *pretends*. These assumptions underlie a model of 'theory of mind' acquisition developed over the years by Leslie and colleagues (Baron-Cohen, 1995; Frith, 1989; German & Leslie, unpublished a, b; Leslie, 1987, 1988, 1991, 1994a,b; Leslie & German, 1995; Leslie & Roth, 1993; Leslie & Thaiss, 1992; Roth & Leslie, 1991, 1998). A central postulate of this model is the existence of the Theory of Mind Mechanism (ToMM), a brain mechanism specialised for attending to mental states. The theory of ToMM is a two factor theory, encompassing a critical distinction between mechanisms underlying the child's conceptual competence and the performance demands that are made by particular 'theory of mind' problems, such as the false belief task.

Competence

As far as competence is concerned, ToMM equips the child with an innate, specialised representational system that employs the basic attitude concepts, BELIEF, DESIRE and PRETEND. These attitude concepts feature in representational structures called 'metarepresentations' (or 'M-representations'), which provide the cognitive system with agent-centred descriptions of behaviour, additionally specifying an *agent*, a 'propositional' *content* and an *anchor* to an aspect of the real world. Two well worn examples of M-representations appear in (1) and (2):

(1) Mother PRETENDS (that) "it is a telephone" (is true of) the banana
(2) Sally BELIEVES (that) "it is in the basket" (is true of) the marble

In example (1), Mother is the agent, who holds the attitude of pretending toward the truth of the proposition 'it is a telephone', while her attitude is anchored in a particular banana in the world. In (2), the agent (Sally) is represented as holding the attitude of believing toward the truth of the content ("it is in the basket") with respect to a particular object (the marble).

The ToMM theory does not, however, claim that the child is equipped with a 'theory' of either pretending or believing. Nor is the child expected to be highly proficient immediately in deploying his conceptual competence when reasoning about these mental states. ToMM provides only the *basis* for acquiring knowledge about mental states. The initial concepts supplied by ToMM allow the child, at least on some occasions, to interpret behaviour in terms of these underlying mental states; the child, by attending to instances of mental states, can begin to learn particular facts about them. Note that this contrasts sharply with the main claim of the conceptual change view, namely, that the meaning of concepts such as BELIEF and DESIRE is given by a 'theory' to which the child subscribes or grasps. On the ToMM view, it is an open empirical question whether or not the child has some innate 'theory' of this or that mental state; but

no such 'theory' is required to ground the M-representation. The child can, for example, simply recognise instances of pretending without knowing what pretending *really is*.

Performance factors in theory of mind tasks

ToMM is initially limited in its effectiveness when 'theory of mind' problems require non-default answers (Leslie, 1991; Leslie & Roth, 1993; Leslie & Thaiss, 1992; Roth, 1993; Roth & Leslie, 1998). Beliefs, by their nature, are normatively true; their contents ought to match reality because that is what beliefs are for (Dennett, 1981; Fodor, 1992; Leslie, 1994a; Leslie & Thaiss, 1992). By default, a belief reasoner should attribute beliefs with contents that are true. In false belief tasks, the default must be over-ridden, and the correct non-factual content must be selected instead. According to the ToMM theory, over-riding the default requires inhibition followed by identification or selection of an alternate content. This inhibition/selection process is handled by a mechanism that co-operates closely with or is part of ToMM, and which has been called the Selection Processor (SP). The SP mechanism develops gradually, becoming increasingly effective over the pre-school years, and beyond. Its gradual increase in efficiency leads to a gradual extension of the child's ability to reason about beliefs, making it progressively easier for children to attend to beliefs that have false contents. In this way, selection processing leverages the child's initial competence.

There is increasing evidence, collected by researchers working within the framework of providing a *task analysis* of 'theory of mind' problems, that information processing demands play an important role in determining the likelihood that children at a given age (and therefore with SP abilities at a given level) will perform well at tasks assessing mental state reasoning (Freeman & Lacohée, 1995; German, 1995a, b; German & Leslie, 1999a, b; Leslie & Polizzi, 1998; Mitchell & Lacohée, 1991; Moore et al., 1995; Riggs, Peterson, Robinson, & Mitchell, 1998; Roth & Leslie, 1991, 1998; Saltmarsh, Mitchell, & Robinson, 1995; Surian & Leslie, 1999; Wellman & Bartsch, 1988; Zaitchik, 1991). In the following sections, we outline some of this evidence consistent with a two factor approach to the development of theory of mind problem-solving ability.

Reasoning about pretend and desire. The flip side to the idea that some 'theory of mind' problems make processing demands that the child cannot meet is the idea that other 'theory of mind' problems make trivial demands. Consider a typical pretend play scenario. Leslie and Happé (1989; see also Leslie, 1987, 1988) argue that children are supported in their task of reading the contents of the protagonist's mental states by the efforts made, on the part of the participants in pretence, to communicate those pretend contents. Shared pretend play is always accompanied by specific communicative displays: 'knowing' looks and smiles, exaggerated intonations, gestures, and actions, explicit verbalisations,

and so on (Leslie, 1987; Piaget, 1962). Success at inferring the content of an agent's pretend, and indeed inferring *that the agent is pretending*, may depend on this communicative display.

Perner et al. (1994), arguing that children are utilising the concept 'PRELIEF' rather than BELIEF and PRETEND, found that 3-year-old children were unable to distinguish whether an actor, putting food in a rabbit hutch (when no rabbit was there), was doing so because the actor *pretended* it or *really believed* it until they were aged about four. However, children in this study were deprived of the cues they might normally use in order to recognise the communicative act of pretending; the character behaves precisely the same way in both the 'pretend' and 'false belief' versions of the scenarios. Therefore, the inference required (from the protagonist's knowledge), is the same in each case. On the view we have been advocating here, where pretenders take no steps to reduce the difficulty of the inferences required by the young child to infer pretence, it is unsurprising that children might be confused between these two cases in the task presented in Perner et al. (1994). It is doubly unsurprising that children who fail to recognise the false belief in a standard false belief task should also fail to distinguish the (unrecognised) false belief from pretence under the same conditions of testing. The only chance a 3-year-old has of doing this is in a 'modified' false belief task where they enjoy greater success with identifying the false belief. Freeman and Lacohée (1995) using a modified 'Smarties' task found 3-year-olds who successfully identified their own false belief were also able to distinguish it from pretending. Interestingly, those children who failed even the modified false belief task were more likely to judge that their claimed knowledge was in fact pretence, suggesting that these children had some inkling that their claim to knowledge was suspect.

An information processing perspective can also account for the early emergence of children's reasoning with the concept desire relative to the concept belief. There is no need to assume, as the theory-theory does, that DESIRE is recognised as a concept in the 2-year-old's innate 'theory of mind' whereas BELIEF is a concept in a later discovered 'theory'. Considering the tasks typically used in assessing understanding of desire versus belief, one can identify task structural differences. While the contents of desire need to be inferred just as do the contents of belief, there is no 'default' content-type that needs to be inhibited prior to the selection procedure. Therefore reasoning about differing desires should stress the SP mechanisms considerably less than reasoning about non-default beliefs. However, might it be possible to arrange circumstances such that an inhibition requirement is introduced into a desire reasoning task?

Moore et al. (1995) did exactly this, by arranging a desire task that closely matched the structure of the false belief task. Children were playing a game with a competitor and initially, both competitors required a certain event to occur to progress (say, that a card turned over should be red)—thus both child and competitor *wanted a red card*. Events were arranged such that children

themselves achieved this interim goal and now could only progress if the card turned over were to be blue, while the competitor was left requiring a red card. Now the child's desire (blue card) has changed, while the competitor's desire remains the same (red card). Children were simply asked what colour card the competitor wanted next. Note that this situation is analogous to the false belief task, where the child and the protagonist start out with one belief (marble in the basket) and the child's belief (but not the protagonist's) is updated (marble in the box). The results showed that 3-year-old children failed this task, predicting that the competitor would have the same desire as the child. Performance was in fact comparable to performance on the standard false belief task. This result is compatible then, with the view that differences in performance in children's reasoning about different mental states can be accounted for in terms of differing information processing demands associated with different task structures (Leslie, 1994a).

Leslie and Polizzi (1998) used a different method of creating an inhibitory process with desire reasoning. Following an idea of Cassidy's (1995), Leslie and Polizzi used a desire that demands what they called 'target-shifting'. The task is to figure which of two locations a protagonist wants when they want the location that does *not* have property X. To identify the NOT(X) location, one must first identify the X location then shift to the alternate. Leslie and Polizzi hypothesised that the brain would have to inhibit its attention to the X location in order to disengage from it and shift to the alternate target. A similar process is hypothesised to occur in selection processing during a false belief task. First, by default, the brain attributes a true belief to a protagonist and thereby identifies a target of belief, then it inhibits attention to that default target and shifts to the alternate target identified by the non-factual content of the false belief. Leslie and Polizzi used a target-shifting desire to probe the existence of the postulated target-shifting (i.e. false) belief. The reasoning was that the combination of two target-shifts would produce a very difficult interaction in which the brain has to return to a previously inhibited target. In one scenario, a sick kitten is in box A. Sally does not want to give her fish to the sick kitten in case she makes it worse. When Sally goes out of the room, the kitten moves from box A to box B. Sally then comes back to her fish. In addition to control questions, the child is asked two standard questions, *Think*: 'Where does Sally think the kitten is?', and Prediction: 'Where will Sally put the fish?'. A group of 4-year-old children who were 100% correct on the Think question, were only 38% correct on the Prediction question.

What makes this last result particularly interesting is, first, performance of 4-year-olds on a target-shifting desire coupled with a *true* belief was near ceiling and, second, all the children who failed prediction in the false belief task had just seconds earlier passed the *Think* question. So the result cannot be accounted for simply by assuming (a) that solving target-shifting desire is hard for 4-year-olds (it is not), and (b) because 4-year-olds are barely able to pass false belief,

any further difficulty pushes them over the edge (these children had *already* solved the false belief and only had to remember the solution for a second or two). So 'vanilla' information processing assumptions will not account for this finding. It needs an interacting-inhibitions SP model such as those discussed by Leslie and Polizzi (1998). Even when the false belief calculation is over, the default belief target remains inhibited, making it difficult for the brain to return to when required by a target-shifting desire.

Reducing difficulty with false belief. Information processing factors play an important role in explaining performance on the false belief task. Several investigations have shown that younger children's performance at the false belief task can be improved by modifying task structure. The SP theory makes specific predictions about the kinds of manipulations that ought to work. We review a few examples here.

One information processing variable assumed to be handled by SP is the task of inhibiting the default attribution of belief contents that reflect the current (true) state of affairs so that the typical error in false belief tasks can be avoided. Consequently, if the need for inhibition is reduced, for example by reducing the salience of the to be inhibited content, children ought to fare better at the task. Zaitchik (1991) and Wellman and Bartsch (1988) employ procedures of this kind. In Zaitchik (1991), children in an 'evidence' condition, who do not get to see that the object has changed location, but are merely told that it has, outperform children given a standard task. The need for SP is lessened here because children are not sure, in the evidence condition, which content is true. In the 'Not Own' belief task (Wellman & Bartsch, 1988), children are told about two possible locations for an object, and asked which location they think is correct. Children are told that a protagonist believes the object to be in the other location, and are asked where the protagonist will look. Again, presumably because children are not sure which location is correct, the need for inhibition is reduced and children pass.

One possible objection that theory-theorists might make to these findings is that children may succeed because they are not required to attribute a belief that they know to be false. German (1995b), addressed this issue using a version of the 'disappear' false belief task (Wimmer & Perner, 1983: Fodor, 1992). In the original disappear condition, the object, rather than being displaced to a second location, is destroyed so there is no 'true' location at which the protagonist can search. Wimmer and Perner's (1983) attempt at this manipulation was partially successful, but ran into the problem that children were unconvinced by the disappear event; Wimmer and Perner had simply removed the bait and placed it 'behind the scenes', with the result that children predicted that the character would look for the supposedly 'destroyed' object behind the scenes. German (1995b) showed in one experiment that destroying the object (chocolate) by eating it also improves performance, but only marginally, because even here

children tend to predict search in the Experimenter's mouth or tummy, probably owing to their limited understanding of digestion (Carey, 1985). In a second experiment, children were presented with disappear conditions presented via cartoon pictures, in order to overcome the problem of destroying an object in real space. Children performed significantly better when presented with a story in which the object is carried off 'far away' by a dog. Although the children do not know exactly where the object is, they do know it is *not* where the protagonist thinks it is.

One potential problem with this finding is that children might succeed by 'guessing' on the basis of an association between the object and the initial hiding place, without taking into account the agent's belief (see for example comments in Perner, 1995, p. 259). A '*seeing*' control (Leslie, 1994a), where the protagonist sees the transfer and therefore should not look in the empty box is one possible check of the 'association strategy'—if this strategy were used then children should also point to the empty box in the *seeing* condition. But there are pragmatic problems in asking children to predict where a character, who has seen an object destroyed, will look for it. Instead, German (1995b), ran additional conditions in which the object was transferred to an intervening location before it disappeared. Children adopting a 'guessing' strategy should split their responses equally between the two locations visited by the object. Instead, children answered correctly on the basis of the protagonists' belief.

A key point as far as these studies is concerned is that unlike Wellman and Bartsch (1988) and Zaitchik (1991), children were required to attribute a *belief they knew to be false*. Reducing the load on the mechanisms responsible for inhibition of the true belief content thus can lead to improved performance in the false belief task.

Different processing routes for Self and Other in the false belief task? Recently, we have begun to explore whether different processing routes are employed when children reason about their own, rather than another person's mental states (German, 1995a; German & Leslie, 1999b). Within the competence-performance framework, it is plausible that reasoning about Self and about Other in 'theory of mind' tasks may differ. For example, a belief held for Self might be explicitly recalled at a later time, provided certain cues to recall are available, whereas for Other, direct recall of the belief contents is impossible; the past belief must be reconstructed from information about the timing of events and the Other individual's exposure to them. Conceptual shift explanations of the acquisition of 'theory of mind', by contrast, have tended to argue that there should be no differences between reasoning about Self and reasoning about Other. The child's performance is constrained by conceptual factors, not how information is processed; understanding the representational nature of belief provides the difficulty in false belief tasks, irrespective of whether the belief is one's own or that of another individual (Gopnik, 1993; Gopnik &

Astington, 1988; Wimmer & Hartl, 1991). Much of the evidence assessing the relative difficulty of belief reasoning for Self versus Other, collected using the deceptive box paradigm (or 'Smarties' task; Hogrefe, Wimmer & Perner, 1986; Perner et al., 1987) has indeed supported this view (Gopnik & Astington, 1988; Riggs & Robinson, 1995a, b; Wimmer & Hartl, 1991).

However, there are good reasons to suppose that the deceptive box task may not be best suited to discovering Self–Other differences should they exist. The false belief that the child is required to identify has been entertained by the child for a very short time and therefore is likely to provide a very small 'target' for direct recall. Moreover, there is no reason why the child at that time should have represented the belief she had as a belief she is having; therefore when asked even a few moments later, after the original belief has been corrected, the child might instead have to reconstruct what belief she should have had under the then prevailing circumstances. The processing route for calculating belief content for Self in this task is thus very similar to the task of calculating the content of the belief that another person should have under those circumstances (see for example Wimmer & Hartl, 1991, p. 127), and is likely to make similar demands on SP.

German (German, 1995a; German & Leslie, 1999b) removed these problems by asking children to explain their own incorrect search action. In one experiment, 3-year-old children (mean 3–6) were required to play the part of 'Sally' in a false belief task. Children were videotaped as they were given a bait object, asked to hide it, and then left the scene long enough for a confederate to move it to a second location. On their return, they searched the wrong location. Children were then shown the video recordings of either themselves, or of another child, going through this procedure. At the critical moment when the incorrect search took place, the video recording was paused and children were asked to explain their own (or the other child's) action, and questioned about their belief at the time of the search. Note that the structure of this task provides children who were asked about their own belief with the opportunity of 'directly' recalling the belief content, which has been entertained for minutes rather than seconds (as in the standard deceptive box task), rendering that content more salient. By contrast, children asked about the belief of another person would still be required to reconstruct the content from the available evidence, with the usual demands on SP.[5] The results showed that children asked about Self were more likely to offer false belief based explanations (46%) and more likely to attribute to themselves a false belief (88%) than were children asked about the beliefs of another (12% and 42% respectively).[6]

These results were replicated and extended in a second experiment, which showed that the Self–Other effect could be obtained without the use of 'video evidence' of the incorrect search, by making use of a different deception procedure. Children themselves transferred a target object (a sticker) from one box to another, before the boxes were both hidden. An intervening task was then administered before a pair of identical 'clone' boxes were removed from the

hiding place, such that when the child searched the box identical to the resting place of the sticker, they were incorrect. Children again were more likely to explain their own incorrect search in terms of their false belief (60%), and were more likely to remember their incorrect belief (80%), than were children asked about another agent (a hand puppet) enacting the same incorrect search (25% and 30% respectively).

These results suggest that children's performance when asked about their own past false beliefs can exceed the performance observed when they are asked about the false beliefs of another. The conceptual deficit view cannot explain these findings because there is no *conceptual* difference between false belief for Self and false belief for Other (Gopnik, 1993; Gopnik & Astington, 1988; Wimmer & Hartl, 1991). However, children are able to recall their own previous belief even when false, if it is made salient enough; otherwise, they must reconstruct the belief from the prevailing circumstances, with the usual selection processing problems. Obviously, only the latter route is available for Other. Standard deceptive box procedures do not seem to allow this difference in strategy, perhaps because the belief is held so transiently (Morton, 1997). There is other evidence that retrieval variables affect the ease with which children succeed at the deceptive box task for their own belief (Freeman & Lacohée, 1995; Mitchell & Lacohée, 1991). Interestingly, we know of no evidence that the same memory procedures help children calculate the beliefs of others. This is consistent with the view that these manipulations help by bolstering a *recall* strategy for Self; where no recall strategy is available (as in the case of Other), these memory manipulations will not work.

Task demands outside the context of mental state reasoning. Also consistent with the performance perspective on children's reasoning and the mind are recent results suggesting that tasks with *no mental state content at all*, but which involve inhibition /selection demands, are problematic for children to the same extent as are tasks involving the attribution of false belief.

Riggs et al. (1998) show that 3-year-old children fail a task where they are required to make an inference on the basis of a counterfactual state of affairs. Children were shown, for example, that round things were to be sorted into one pot, while long and thin things were to be placed into another pot. Children were shown a round lump of dough and it was put into the 'round things' pot. The experimenter then took the dough, squashed it into a long and thin shape and asked the children where it should go now. It was placed in the 'long things' pot and children were then asked where the object would be if the experimenter hadn't played with it. Note that this task requires children to inhibit the current shape of the object and select a counterfactual alternative description in order to pass the task. They performed at levels similar to their performance at false belief, and performance on the two tasks was also correlated, suggesting common performance demands.

Roth and Leslie (1998) provide another example of a task that 3-year-olds find difficult, despite its involving no mental state content. This task, called the 'screen task' was designed to match the Sally-Anne task in important structural details. Children are presented with an opaque screen. In front of this screen are placed a box and a basket, and a marble is placed inside the basket. All these items are then moved behind the opaque screen, and an identical box and basket are placed in corresponding positions in front of the screen. An identical marble is placed in the basket in front of the screen, left there for a few moments and subsequently moved to the box in front of the screen. The child is asked: 'Behind the screen, where is the marble?'. This task, it was hypothesised, would stress SP to some extent; children were faced with a question (about the situation behind the screen) that can be answered only by inhibiting the similar and salient state of affairs in front of the screen. Four-year-old children performed similarly on the screen task and a standard false belief task. Younger children, by contrast, were found to perform poorly at both the screen and standard false belief tasks, and the two tasks were correlated.

CONCLUSIONS

We have outlined two main approaches to the puzzle of how we acquire a 'theory of mind'. We have criticised theory-theory approaches that assume that children 'discover' a set of 'theories' about the nature of mental states. Children's performance on 'theory of mind' tasks is attributed to the sophistication of the 'theories' they possess. Children fail the false belief task, for example, when their 'theory' of belief does not stipulate that beliefs are representations. Theory-theories have yet to address the information processing questions that their claims raise, for example, how new concepts and 'theories' are devised by young, very ordinary children, how 'evidence' is identified, and how 'environmental interaction' leads to specific abstract formulations.

The ToMM competence theory suggests that children are equipped with a set of innate concepts that simply allow children to attend to mental states, and thus to learn about them. Concept possession is not assumed to depend upon possession of a package of knowledge or 'theory', and mental state concepts are not assumed to depend upon a package of knowledge other than the M-representation itself. Children are not initially expert in deploying their conceptual repertoire, their task is just beginning. To become effective reasoners about mental states, they need to draw on problem solving capacities that develop only gradually over the pre-school period. Children's proficiency at reasoning about pretending and desiring is likely to be influenced by the relatively low performance demands inherent in tasks assessing these mental states, while the standard false belief task is solved rather later, after children develop the resources to deal effectively with the inhibition and selection difficulties involved in these tasks. Only through carefully teasing apart competence and performance issues, will progress be made in understanding 'theory of mind' development.

NOTES

1. We use uppercase when we refer to concepts and lower case when we refer to the property 'in the world' that the concept designates. Thus, DOG is our mental representation, or concept, that designates the property *being a dog*. One of the potentially confusing things about 'theory of mind' is that both the concepts and the properties they designate are 'in the mind'. While on notation, we try to remember to place the 'theories' that are attributed by scientists to children in scare quotes to distinguish them from the theories that scientists devise: so, for example, Gopnik has devised a theory that children have 'theories'.
2. Gopnik (1996) has also argued for the symmetrical view of 'the scientist as child'.
3. Perner (1995) identifies this distinction with one made by Goodman (1976), and retreats from his (1991, pp. 19–20) terminology where he claimed that children must grasp the distinction between *sense* and *referent* (Perner, 1995, footnote 5).
4. Perner borrows this use of the term 'situation' from Barwise and Perry (1983) where it means roughly what others mean by the term 'proposition' – something that can be true or false—as opposed to what is usually meant by the term, a (non-intentional) state of affairs. Here, we use 'situation' and 'proposition' interchangeably.
5. It might be argued that the availability of evidence of incorrect search might help the child to reason 'back' to the false belief, but Moses and Flavell (1990) provide data suggesting that such extra evidence (i) only helps with belief attribution at the expense of failed control questions and (ii) does not result in a high proportion of belief based explanations for another's incorrect search (≈ 12%).
6. The belief based explanation referred variously to the child's lack of knowledge, lack of perceptual access to the transfer events, memory of leaving the object in the initial location, as well as explicit mention of false belief.

ACKNOWLEDGEMENTS

We would like to thank Peter Mitchell and Kevin Riggs for helpful comments on an earlier version of this chapter.

REFERENCES

Baron-Cohen, S. (1995). *Mindblindness*. Cambridge, MA: Bradford Books, MIT Press.

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

Barwise, J., & Perry, J. (1983). *Situations and attitudes*. Cambridge, MA: MIT Press.

Campione, J., Brown, A, & Ferrera, R. (1982). Mental retardation and intelligence. In R.J. Sternberg (Ed.), *Handbook of human intelligence*. Cambridge: Cambridge University Press.

Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: Bradford Books, MIT Press.

Carey, S. (1988). Conceptual differences between children and adults. *Mind and Language, 3*, 167–183.

Carey, S. (1991). Knowledge acquisition: Enrichment or conceptual change? In S. Carey & R. Gelman (Eds.), *The epigenesis of mind: Essays in biology and cognition*. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Carey, S. (1995). On the origins of causal understanding. In D. Sperber, D. Premack, & A. Premack (Eds.), *Causal cognition*. Oxford: Clarendon Press.

Carey, S., & Spelke, L. (1994). Domain-specific knowledge and conceptual change. In L. Hirschfeld & S. Gelman (Eds.), *Mapping the mind: Domain specificity in culture and cognition*. New York: Cambridge University Press.

Carey, S., & Spelke, L. (1996). Science and core knowledge. *Philosophy of Science, 63*, 515–533.

Cassidy, K.W. (1995, April). *Use of a desire heuristic in a theory of mind task*. Poster presented at the Biennial Meeting of the Society for Research in Child Development, Indianapolis, USA.

Chi, M.T.H. (1992). Conceptual change within and across ontological categories: Examples from learning and discovery in science. In R. Giere (Ed.), *Cognitive models of science: Minnesota studies in the philosophy of science*. Minneapolis, MN: University of Minnesota Press.

Dennett, D.C. (1981). Three kinds of intentional psychology. In R. Healey (Ed.), *Reduction, time and reality*, (pp. 37–61). Cambridge: Cambridge University Press.

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

Forguson, L., & Gopnik, A. (1988). The ontogeny of common sense. In J. Astington, P.L. Harris, & D. Olson (Eds.), *Developing theories of mind* (pp. 226–243). Cambridge: Cambridge University Press.

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

Frith, U. (1989). *Autism: Explaining the enigma*. Oxford: Blackwell.

Gelman, R. (1991). Epigenetic foundations of knowledge structures: Initial and transcendent constructions. In S. Carey & R. Gelman (Eds.), *The epigenesis of mind: Essays in biology and cognition*. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

German, T.P. (1995a). *Children's explanation of action: Desires versus beliefs in theory of mind*. Unpublished Ph.D. thesis, University of London.

German, T.P. (1995b, April). *Testing Fodor's theory of the child's theory of mind*. Poster presented at the biennial meeting of the Society for Research in Child Development, Indianapolis, IN, March, 1995.

German, T.P., & Leslie, A.M. (1999a). Children's inferences from KNOWING to PRETENDING and THINKING. Manuscript, Department of Psychology, University of Essex. Paper submitted for publication.

German, T.P., & Leslie, A.M. (1999b). Self–other differences in false belief: Recall versus reconstruction. Manuscript, Department of Psychology, University of Essex. Paper submitted for publication.

Goodman, N. (1976). *Languages of art*. Indianapolis, IN: Hackett Publishing Co.

Gopnik, A. (1993a). How we know our minds: The illusion of first-person knowledge of intentionality. *Behavioral and Brain Sciences, 16*, 1–14.

Gopnik, A. (1993b). Theories and illusions. *Behavioral and Brain Sciences, 16*, 90–100.

Gopnik, A. (1996a). Theories and modules; creation myths, developmental realities, and Neurath's boat. In P. Carruthers & P.K. Smith (Eds.). *Theories of theories of mind* (pp. 169–183). Cambridge, Cambridge University Press.

Gopnik, A. (1996b). The scientist as child. *Philosophy of Science, 63*, 485–514.

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

Gopnik, A., & Meltzoff, A. (1997). *Words, thoughts and theories*. Cambridge, MA: MIT Press.

Gopnik, A., & Wellman, H. (1992). Why the child's theory of mind really *is* a theory. *Mind and Language, 7*, 145–171.

Gopnik, A., & Wellman, H. (1994). The 'theory theory'. In L. Hirschfeld & S. Gelman (Eds.), *Mapping the mind: Domain specificity in culture and cognition* (pp. 257–293). New York: Cambridge University Press.

Hadwin, J., & Perner, J. (1991). Pleased and surprised: Children's cognitive theory of emotion. *British Journal of Developmental Psychology, 9*, 215–234.

Harris, P.L. (1996). Desires, beliefs and language. In P. Carruthers & P.K. Smith (Eds.), *Theories of theories of mind* (pp. 200–220). Cambridge, Cambridge University Press.

Hogrefe, G., Wimmer, H., & Perner, J. (1986). Ignorance versus false belief: A developmental lag in attribution of epistemic states. *Child Development, 57*, 567–582.

Johnson, S., & Carey, S. (1998). Knowledge enrichment and conceptual change in folkbiology: Evidence from Williams Syndrome. *Cognitive Psychology, 37*, 156–200.

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

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

Leslie, A.M. (1991). The theory of mind impairment in autism: Evidence for a modular mechanism of development? In A. Whiten (Ed.), *Natural theories of mind: Evolution, development and simulation of everyday mindreading* (pp. 63–78). Oxford: Blackwell.

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

Leslie, A.M. (1994b). ToMM, ToBy and agency: Core architecture and domain specificity. In L. Hirschfeld & S. Gelman (Eds.), *Mapping the mind: Domain specificity in culture and cognition* (pp. 119–148). New York: Cambridge University Press.

Leslie, A.M., & German, T.P. (1995). Knowledge and ability in 'theory of mind': One-eyed overview of a debate. In T. Stone, & M. Davies (Eds.), *Mental simulation: Evaluations and applications* (pp. 123–150). Oxford: Blackwell.

Leslie, A.M., & Happé, F.G. (1993). Autism and ostensive communication: The relevance of metarepresentation. *Development and Psychopathology, 1*, 205–212.

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

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.

Lillard, A.S. (1993). Young children's conceptualization of pretense: Action or mental representational state? *Child Development, 64*, 372–386.

Lillard, A.S. (1996). Body or mind: Children's categorization of pretense. *Child Development, 64*, 348–371.

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

Moore, C., Jarrold, C., Russell, J., Lumb, A., Sapp, F., & MacCallum, F. (1995). Conflicting desire and the child's theory of mind. *Cognitive Development, 10*, 467–482.

Morton, J. (1997). Free associations with EPS and memory. *The Quarterly Journal of Experimental Psychology, 50A*, 924–941.

Moses, L.J., & Flavell, J.H. (1990). Inferring false beliefs from actions and reactions. *Child Development, 61*, 929–945.

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.

Perner, J. (1993). The theory of mind deficit in autism: Rethinking the metarepresentational theory. In S. Baron-Cohen, H. Tager-Flusberg, & D. Cohen (Eds.), *Understanding other minds: Perspectives from autism* (pp. 112–137). Oxford: Oxford University Press.

Perner, J. (1995). The many faces of belief: Reflections on Fodor's and the child's theory of mind. *Cognition, 57*, 241–269.

Perner, J., Baker, S., & Hutton, D. (1994). Prelief: the conceptual origins of belief and pretense. In C. Lewis & P. Mitchell (Eds.), *Children's early understanding of mind: Origins and development* (pp. 261–286). Hove, UK: Lawrence Erlbaum Associates Ltd.

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

Piaget, J. (1929). *The child's conception of the world*. London: Routledge & Kegan Paul.

Piaget, J. (1962). *Play, dreams and imitation in childhood*. London: Routledge & Kegan Paul.

Pinker, S. (1997). *How the mind works*. New York: Norton.

Riggs, K.J., & Robinson, E.J. (1995a). What people say and what they think: Children's judgements of false belief in relation to their recall of false messages. *British Journal of Developmental Psychology, 13*, 271–284.

Riggs, K.J., & Robinson, E.J. (1995b). Children's memory for actions based on a false belief. *Journal of Experimental Child Psychology, 60*, 229–244.

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

Roth, D. (1993). *Beliefs about false beliefs: Understanding mental states in normal and abnormal development*. Unpublished 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.

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

Saltmarsh, R., Mitchell, P., & Robinson, E.J. (1995). Realism and children's early grasp of mental representation: Belief-based judgements in the state-change task. *Cognition, 57*, 297–325.

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

Vosniadu, S., & Brewer, W.F. (1992). Mental models of the Earth: A study of conceptual change in childhood. *Cognitive Psychology, 24*, 535–585.

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

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

Wellman, H.M., & Woolley, J.D. (1990). From simple desires to everyday beliefs: The early development of everyday psychology. *Cognition, 35*, 245–275.

Wimmer, H., & Hartl, M. (1991). Against the Cartesian view on mind: Young children's difficulty with own false belief. *British Journal of Developmental Psychology, 9*, 125–138.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: representation and constraining function of wrong beliefs in children's understanding of deception. *Cognition, 53*, 45–57.

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

CHAPTER TWELVE

Children's understanding of belief: Why is it important tc understand what happened?

Heinz Wimmer and Andreas Gschaider
University of Salzburg, Austria

In this chapter we try to show that the mature concept of belief entails more tha what usually is examined in developmental "theory of mind" studies. In partic lar we will argue that the concept of belief entails that the wrong thought whi we took to be reality was caused in such a way that left no possibility other tha to take it to be reality. In other words, we not only understand what we wrong believed to be the case, but also that we did so for good reasons. In the case false beliefs about simply situational facts (like what something is or where is), these good reasons are important as they save our view of ourselves healthy persons who are not hallucinating, and they also save our view of orderly world where objects do not change their identity or their place in unpr dictable ways. Understanding the causation of wrong thoughts about what the case is also responsible for the fact that at times we find false beliefs funny or tragic.

For the first author the issue of children's understanding of the causation knowledge and belief has some tradition and what is proposed here is actually reformulation of the theoretical stance originally formulated more than 10 yea ago in Wimmer, Hogrefe, and Sodian (1988). The old formulation was that f mastering the false belief tasks developed in Salzburg (Hogrefe, Wimmer, Perner, 1986; Wimmer & Perner, 1983), the child has to infer what anoth person believes from the misleading informational circumstances the other pe son was exposed to. The main hypothesis was that children below the age of years or so lack the conceptual basis for the inference, namely an understandi of the causal relationship between informational circumstances on the one ha and resulting epistemic states on the other.