Natural Theories of Mind
A. Whiten (ed)
Blackwell (1991)
Oxford UK

2

From Desires to Beliefs: Acquisition of a Theory of Mind

HENRY M. WELLMAN

The topic of this volume concerns the emergence of an understanding of mind. The term 'mind', as I am using it, is an everyday one rather than a scientific one, it is a term within an everyday theory. This everyday theory construes overt human behaviour as the consequence of covert mental states, such as the actor's beliefs, hopes, ideas, and desires (see also chapter 1). When do children adopt this everyday mentalistic construal of human behaviour?

This question has a complicated and fascinating answer. To be brief, however, I am going to answer simply: at three years. To elaborate on this simplistic answer I discuss three points. First, I will sketch what I believe constitutes our everyday theory of mind. This analysis revolves around characterizing everyday mentalism as a belief–desire psychology. Secondly, I review some recent findings showing that children as young as three years understand and utilize belief–desire psychology and hence evidence a first theory of mind. Other chapters in this volume and elsewhere (e.g. Astington, Harris and Olson, 1988) review related research; I will concentrate on the research of my collaborators and myself. Thirdly, I suggest that children younger than three, say two-year olds, fail to understand belief–desire psychology. They utilize instead, a simple desire psychology.

Sketch of a Theory of Mind

There are two intuitive aspects of our understanding of the mental states of self and other. Crudely put, I will call these the hypothetical and causal

Support for this research was provided by grant HD–22149 from NICHD.

aspects of mind. The essence of the hypothetical aspect is our understanding of the difference between thoughts or ideas on the one hand and objects or overt behaviours on the other. For example, a thought about a dog is not the same sort of thing as a dog. Indeed, a thought about a dog is not the same sort of thing as even a shadow of a dog, or a photograph of a dog. Mental entities are internal, subjective, non-real and hypothetical, whereas physical entities are real, external, substantial and objective. The phrase *theory of mind*, as a label for our everyday theory, emphasizes our ordinary understanding of this hypothetical nature of mind. My collaborators and I have researched young children's understanding of the hypothetical, non-real nature of mental entities, and find that even three-year-olds appropriately distinguish mental entities from physical entities (cf., Wellman and Estes, 1986; Estes, Wellman and Woolley, 1989).

According to our everyday theory of mind, our thoughts – our beliefs, plans, ideas and so on – not only are mentalistic but are also causal. A useful and typical short-hand here is to divide our causal mental states into two generic sorts: beliefs and desires (e.g., Davidson, 1980). The causal aspect of mind, then, depicts overt human actions as the joint product of the actor's beliefs and desires. The phrase *belief–desire psychology*, as a label for our everyday theory, highlights this causal aspect of mind. The hypothetical and causal aspects of mind are intimately interrelated. Theory of mind and belief–desire psychology are different descriptive labels for the same topic, the same basic set of understandings. In this chapter, however, I approach the topic from the perspective of the causal aspect of mind.

At the centre of our everyday belief–desire psychology is a basic triad: beliefs, desires, and actions.

(1) Why did Bill go to the swimming pool? He *wanted* to swim and *thought* the pool was open.
(2) Why did Jill watch television? She *wanted* some entertainment and *thought* that a programme she liked was on.

The fundamental, obvious idea is that people engage in actions because they *believe* those actions will satisfy certain *desires*. Of course the theory is more complicated than this in a variety of important ways. It includes reasoning such as:

(3) Why did John go to the candy machine? He was *hungry* and *wanted* a candy bar, and *thought* he'd *seen* the kind he liked in that machine.

Hence, belief–desire psychology incorporates not only beliefs, desires and actions but a network of related constructs such as physiological states (e.g., 'he was hungry') and perceptions (e.g., 'he'd seen that kind'). Figure 2.1 captures the organization of these related constructs. Briefly and simplistically, physiological states and basic emotions ground one's desires. Beliefs, on the other hand, are often derived from perceptual experiences. Furthermore, one's actions lead to outcomes in the world, and these outcomes lead to reactions. As depicted in figure 2.1, at least two basic sorts of reactions are encompassed by the theory: reactions dependent on desires and reactions dependent on beliefs. That is, the outcome of an action can satisfy or fail to satisfy the actor's desires, leading, generically, to happiness reactions. You want something and get it and you are happy, or you fail to get it and you are sad or angry. Also, the outcome of an action can match or fail to match an actor's beliefs, generically termed surprise reactions. You think something will happen and it does not so you are surprised or puzzled.

In this system the key mental states – beliefs and desires – cause actions. Such states are private and not directly observable in others, albeit experienced directly in oneself. However, others' mental states (and at times our own) can be inferred from, among other things, perceptual experiences (e.g., what he sees), from physiological history (e.g., how long it's been since he's eaten) and from emotional expressions and reactions (e.g., when he's happy or when he's surprised).

This brief sketch does not provide a full or proper account of adult belief–desire reasoning. It leaves unelaborated several concepts, such as intention (see chapters 3 and 4) and traits (see Wellman, 1990, chapter 4). But even this crude sketch is sufficient to begin to ask, when do children engage in reasoning of this sort? As is clear from the many connections and links depicted in figure 2.1, if children engage in this sort of reasoning they should know several things. But, most centrally there seem to be two basic sorts of reasoning to assess, that require a direct understanding of the belief, desire, action triad. First, in the forward direction, if children know an actor's beliefs and desires they should be able to predict his behaviour. Secondly, in the backward direction, if children are given an action to explain, they should explain it by appeal to beliefs and desires.

Three-Year-Olds' Belief–Desire Psychology

Predictions

Can three-year-olds appropriately predict behaviour, given information as to an actor's beliefs and desires? In our studies of this sort (Wellman and Bartsch, 1988), we start by telling children about a character who desires

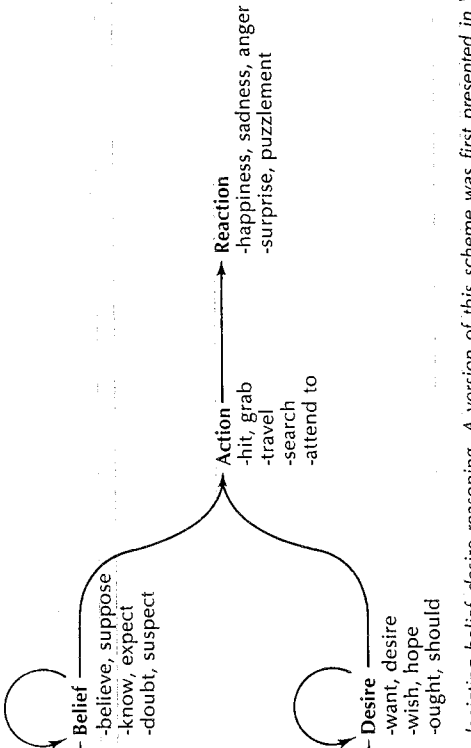

Figure 2.1 *Simplified scheme for depicting belief–desire reasoning. A version of this scheme was first presented in Wellman and Bartsch (1988).*

something – for example, 'Sam wants to find his puppy'. But, 'the puppy is lost, and it might be hiding in the garage *or* under the porch'. Then the child is told about the character's belief, 'Sam thinks his puppy is under the porch', and asked where Sam will look, in the garage or under the porch? This task requires the child to use information about the character's desire coupled with his belief to predict his action. For reasons that will become clear as we go along, the most critical and difficult thing to demonstrate is that young children understand belief. That is, do young children understand that actors not only will be directed towards things that they want, but that their actions will also be constrained by their beliefs? This sort of task, a *Standard Belief* task, attempts to test whether children understand that the actor's beliefs must be taken into account, in addition to the actor's desire, in order to predict action. Since the puppy might be in either location, knowing only Sam's desire leads to the prediction that he could look in either place. But, adding in Sam's belief unambiguously narrows the prediction to a single location.

Unfortunately children could respond appropriately on Standard Belief tasks without really understanding beliefs, that is, by using some less interesting alternative strategy. To control for alternative strategies we have devised a number of variations on this task. For example, in *Not-Belief* tasks children were told 'Sam thinks his puppy is *not* in the garage', so that they could not be correct simply by citing the last location mentioned in the story. Instead, the correct prediction is the unmentioned location.

Suppose by chance that our belief statements about Sam consistently coincide with the subject's own belief – the child thinks puppies will really hide under porches, not in garages. In this case, the subject might be correct not by understanding beliefs, but simply by reporting what she herself would do – 'The puppy is lost, I'd look under the porch'. In *Not-Own Belief* tasks children were first asked about their own beliefs, for example, 'Where do *you* think the puppy is?' After the child stated her belief she was told that the character in fact has the opposite belief.

A troubling possibility is that children might use what we have called *reality assessment strategies* to reason about these problems. Suppose that the young child has no conception of belief and hence does not understand belief statements such as 'Sam thinks his dog is in the garage'. Instead, children misinterpret these statements as specifying the real state of affairs. For example, on Standard Belief tasks, the statement 'Sam thinks his dog is in the garage' is understood as a statement about reality. Then the child would reason, 'Sam wants his dog, it's really in the garage, so Sam will look there'. In *Discrepant Belief* tasks we described to children a situation where there really were targets in two locations. Use of a reality assessment strategy should lead to predicting the character would search in either location or

both (since targets are in both). An understanding of belief should lead to the single correct prediction.

We have conducted tests of children's belief–desire reasoning in ten different conditions such as these, controlling various alternative interpretations (e.g. Wellman and Bartsch, 1988; Wellman, 1990, chapter 3). The four conditions described above exemplify our reasoning. The top portion of figure 2.2 captures the nature of these four different conditions with respect to some competing hypotheses. On the left, if children understand this basic sort of belief–desire causation then they should be uniformly correct across all versions of the task. In the middle there is a depiction of children's responses if they simply predict what they themselves would do. In this case children could be coincidentally correct on the Standard and Not-Belief tasks if their belief consistently coincides with the belief stated for the actor. However, they would be incorrect on Not-Own Belief tasks, and they should respond at chance on Discrepant Belief tasks since on those tasks their own belief is that target objects are in both locations. The right-hand graph depicts what would happen if children make predictions based on reality assessment – predicting the character will look where there really are targets. In this case they could be correct on most tasks but only at chance on Discrepant Belief tasks.

The bottom portion of figure 2.2 shows the data from 20 to 40 three-year-olds in a condition. They are consistently correct across all versions of the task. Even on Discrepant Belief tasks three-year-olds are 82 per cent correct (significantly greater than a chance value of 50 per cent) where they have to predict the other's action on the basis of the other's belief which is discrepant from their own belief and indeed from reality.

Now it is true, as many studies have shown, that on *false* belief prediction tasks three-year-olds perform poorly (see for example Perner et al., 1987). We devised what we think of as the simplest false belief tasks possible, termed Explicit False Belief tasks. In such tasks the child is told of reality and told that the character has a false belief. For example, 'Sam's puppy is in the garage; Sam thinks his puppy is under the porch'. When given tasks of this sort, three-year-olds were only correct 16 per cent of the time. However, I believe that this result shows a peculiar difficulty that young children have with understanding *false* beliefs, it does not indicate a lack of understanding of beliefs altogether. That is, young children understand belief in spite of misunderstanding false beliefs. Children's success in our other tasks demonstrates an understanding of belief, I think, as do their explanations of actions.

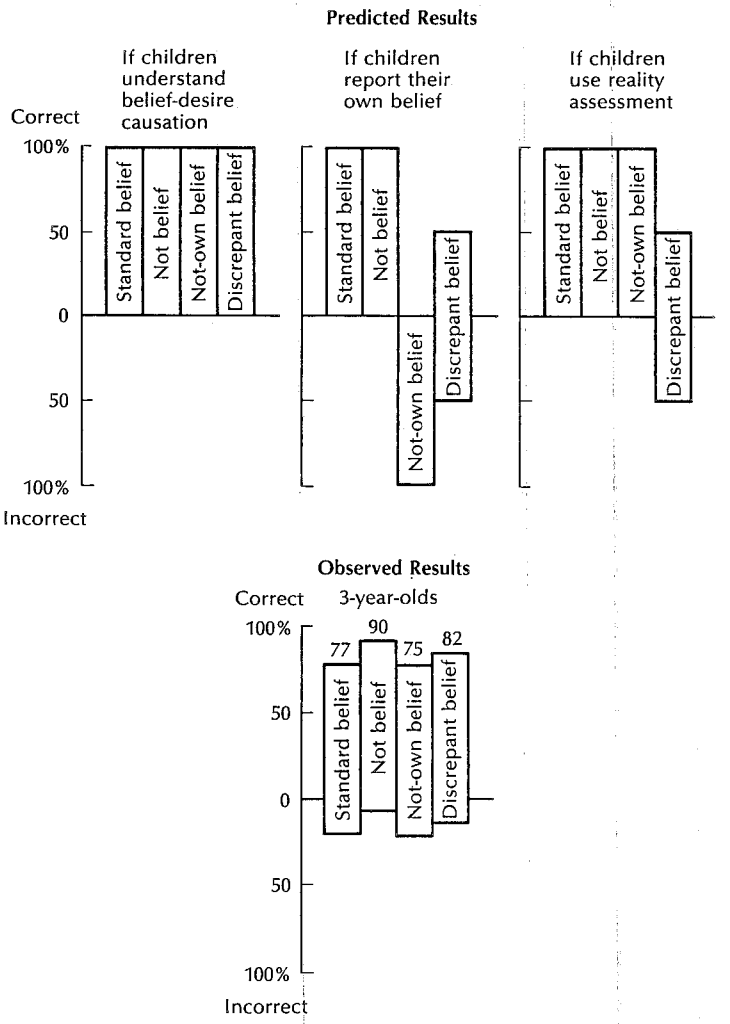

Figure 2.2 *Predicted and observed responses to belief–desire prediction tasks.*

Explanations

Our basic explanation task is this (Bartsch and Wellman, 1989): We tell children of a simple act, for example, 'Jane is looking for her kitten under the piano', and then ask children to explain the act: 'Why do you think she is doing that?' As a look at figure 2.1 suggests, belief–desire psychologists should construct such explanations by appeal, proximally, to the characters' beliefs and desires, and by appeal more generally to beliefs, desires, perceptions, physiology, and basic emotions. Appeal to beliefs and desires narrowly I will call belief–desire explanations; appeal to any of the constructs in the larger scheme – belief, or perception, or emotion, etc. – I will call psychological explanations.

We presented three-year-olds, four-year-olds, and adults with the three different sorts of explanation items shown in box 2.1: Neutral items, Anomalous Desire items (designed to pull more for desire explanations), and Anomalous Belief items (designed to pull more for belief explanations). After each item was presented we asked 'Why do you think the actor is doing that?' And then if children did not spontaneously mention beliefs or desires we mildly prompted them by asking simply, 'What does she want?' or 'What does she think?'

Were the explanations generated of the sort predicted by the theory sketch? That is, were they psychological explanations, generally, and belief–desire explanations specifically? What else could they have been? They could have been behaviouristic explanations, invoking for example a history of conditioning – e.g., 'She's found it there again and again in the past'. Or, they could have been physicalistic explanations – e.g., 'The wind

Box 2.1

Neutral items:	Here's Jane. Jane is looking for her kitten under the piano. Why do you think Jane is doing that?
Anomalous Desire items:	Here's Jane. Jane hates frogs. But Jane is looking for the frog under the piano. Why do you think Jane is doing that?
Anomalous Belief items:	Here's Jane. Jane is looking for her kitten. The kitten is hiding under the chair. But Jane is looking under the piano. Why do you think Jane is doing that?

blew her over there'. Neither children nor adults ever offered behaviouristic or physicalistic sorts of explanations. Both children and adults did, however, make simple referrals to the external situation as in, 'the kitten was lost'. They did so about 30 per cent of the time. But, more often than that children and adults provided psychological explanations, that is, they explicitly referred to the actor's perceptions, beliefs, desires and basic emotions as explanations for his act. As figure 2.3 shows, three-year-olds used psychological explanations 65 per cent of the time in their unprompted explanations; this is not different from the percentage for four-year-olds and adults. Moreover, of three-year-olds' unprompted psychological explanations, 60 per cent referred specifically to beliefs or desires. For comparison the percentages for four-year-olds and for adults were 69 per cent and 67 per cent respectively.

So in response to 'Why did she do that?', three- and four-year-olds and adults give the same general sorts of explanations: psychological explanations broadly, and belief–desire explanations specifically. However, this analysis

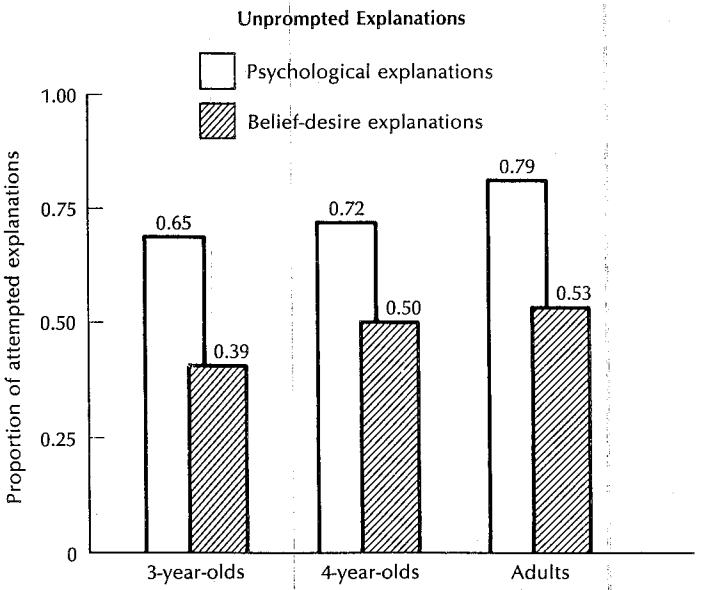

Figure 2.3 *Types of unprompted explanations of actions offered by children and adults.*

considers belief and desire explanations jointly; what about belief explanations alone? As it happens, on most items children and adults initially prefer desire explanations. In our data, at all ages, desire explanations outweighed belief explanations about two to one. It was because of this general preference for desire explanations that our methods included Anomalous Belief tasks.

Subjects received three Anomalous Belief stories like the one shown in box 2.1. Sixty per cent of the children gave an *unprompted* belief explanation to at least one of these three stories. For example, when told 'Jane's kitty is under the chair, but Jane is looking for it under the piano, why is she doing that?', children answered 'She thinks it's under the piano' or 'She doesn't know where it is'. Further, let's suppose that the child did not first give a belief explanation but instead, a desire one such as, 'She wants to find her kitty'. Then he or she was mildly prompted with 'What does Jane think?' When that happened 74 per cent of three-year-olds, 91 per cent of four-year-olds and 100 per cent of adults gave relevant, appropriate belief explanations.

These data on children's explanations of action show, I think, that most children by the age of three years construe human actions as the product of mental states of belief and desire, and can reason about a person's beliefs as well as desires. Let me reiterate the importance of the construct of belief in this reasoning and hence in our everyday theory of mind. Desires motivate behaviours, but beliefs frame them. Persons' actions can thwart their own desires because beliefs are also at work. Remember Jane who wants her kitty which is under the chair, but who looks for it under the piano. Why? Because she *thinks* it's under the piano.

Two-Year-Olds' Simple Desire Psychology

If even three-year-olds possess a belief–desire psychology, maybe infants do. Fodor for example has suggested that we are born with such a construal of human behaviour (Fodor, 1987). However, I propose that three years is just about the earliest age at which children understand belief and thus can participate in belief–desire reasoning.

Let me begin by briefly comparing three broad classes of psychological theory: *behaviourism*, *internal state theory* and *cognitive theory*. Behaviourism, as I mean it, attempts to explain action solely on the basis of functional relations between observable states, specifically stimuli and responses. Behaviourism eschews attributions about the 'insides' of the organism. In contrast, internal state psychologies, such as classic drive theory, endow

organisms with internal states. Drive theory, for example, attributes to organisms internal drives such as hunger, whose waxing and waning propel behaviour. Cognitive theories are, in the broadest sense, internal state theories too. But they constitute a specific refinement of internal state theory, since they invoke certain very distinctive internal *representational* states that function to provide an internal mental world for the organism, such as in everyday terms, a person's beliefs.

What characterizes young children's thinking about people; do they conceive of people in behaviouristic, internal state, or cognitive terms? As reviewed thus far, by three years children are not behaviourists and are not merely internal state psychologists: they are cognitivists. Specifically, theirs is a mentalistic naive psychology, incorporating a rudimentary conception of belief. In contrast, I propose that most two-year-olds are internal state psychologists, although of a peculiar kind. Specifically, theirs is a simple desire psychology which includes no conception of belief.

Let me first elaborate on how a simple desire psychology might be possible – a psychology of action that considers only desires, and importantly considers desires only in a nonrepresentational sense. Figure 2.4 is an attempt graphically to capture this simple conception of desire and to contrast it with a conception of belief. Beliefs are representational; to attribute to someone the thought 'that is an apple' involves construing the other as representing an apple in their mind. Simple desires in the sense I am trying to carve them out here, however, require no attribution to the other of a representation. In this simple understanding, to say that 'he wants an apple' attributes to the other an internal state, for example a state of longing, but an internal state for an *external object*. Simple desires embody no notion of representing an apple in your head, simply *wanting* one. In short, I think it is possible to imagine a simple desire psychology – one resting essentially on a conception of internal states directed towards obtainment of objects in the world – and in this way quite different from a belief–desire psychology which rests centrally if not wholly on a conception of internal cognitive states representing truths about the world.

What sort of reasoning about actions might be encompassed by a simple desire psychology? Essentially, simple desires can cause actors (1) to engage in goal-directed actions, including persisting in certain actions if the goal is blocked, and (2) to have certain emotional reactions (getting what you desire yields happiness, not getting it produces frustration, unhappiness, etc.). Such a desire psychology thus provides some simple but cogent accounts and predictions of various acts. Thus, if a simple desire psychologist knows that 'Sam wants an apple', he can predict that Sam will look for an apple. And if he knows 'Sam wants a specific apple' and that 'the apple is in the kitchen', he can predict that Sam will look in the kitchen. He can predict

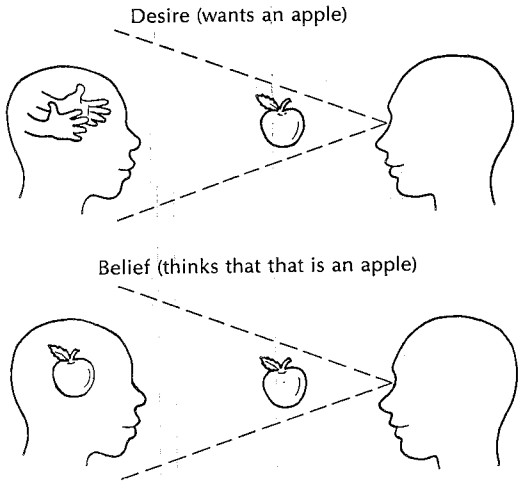

Figure 2.4 *Simple desires (top) and ordinary beliefs (bottom). Taken from Wellman and Woolley (1990).*

that 'Sam will look in the kitchen' under the general maxim that people act to fulfil their desires. A desire psychologist can also predict that if Sam finds the apple he will be happy, under the general maxim that getting what you want makes you happy.

A Study of Two-Year-Olds' Understanding of Desire Psychology

Can two-year-olds reason via a simple desire psychology? Jacqui Woolley and I (Wellman and Woolley, 1990) conducted a preliminary study to find out. As shown in the left-hand portion of figure 2.5, in our study children made judgements about the actions and emotional reactions of story characters in each of three types of situations. In the Finds-Wanted situation a doll character wants something that may be in one of two locations, the character searches in one location and gets the object. The Finds-Nothing situation was identical to Finds-Wanted except that upon searching in the first location nothing was there. The Finds-Substitute situation was identical to Finds-Wanted except that upon searching in the first location the

character found an attractive object, but not the one said to be wanted.

In making *action* judgements (at the top of figure 2.5) children had to predict the character's subsequent action, that is whether he or she would go on to search in the second location or would stop searching. An understanding of the implications of characters' desires should lead to a prediction

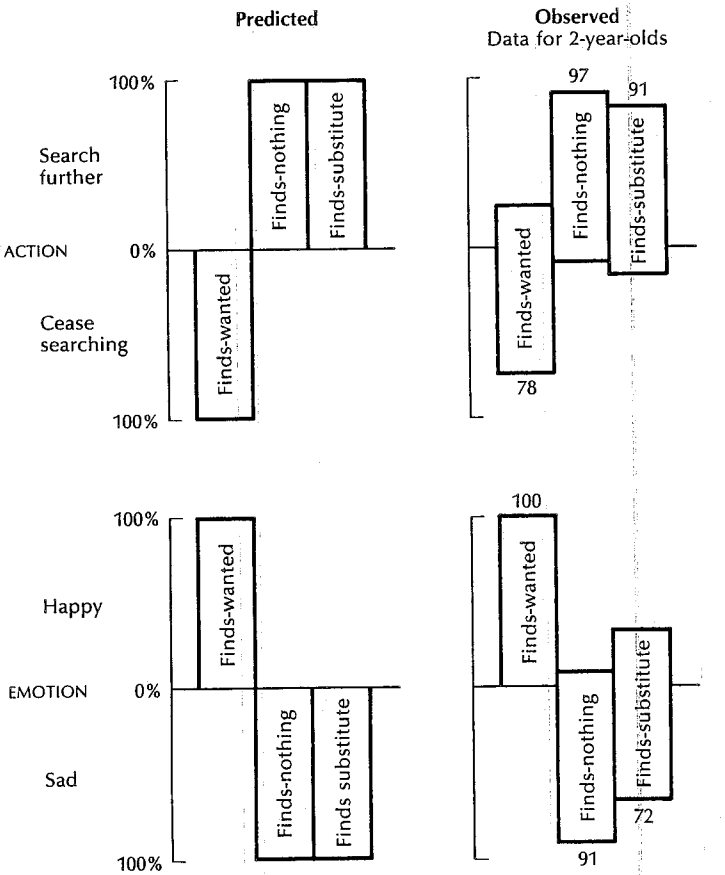

Figure 2.5 *Predicted (left) and observed (right) responses for action and emotion judgements from a study of two-year-olds' understanding of simple desires. Taken from Wellman and Woolley (1990).*

of continued search in the Finds-Nothing and Finds-Substitute situations but to cessation of search in the case of Finds-Wanted. In making *emotion* judgements (at the bottom of figure 2.5) children had to state the character's emotional reaction, whether he or she was happy or sad. An understanding of the role of desires in mediating emotional reactions should yield a prediction of happiness in the Finds-Wanted situation but sadness in the Finds-Nothing and in the Finds-Substitute situations.

So the left-hand portion of figure 2.5 shows the conditions in this study and the predicted patterns of results if children understand the role of desires in predicting actions and emotional reactions. The right-hand portion of that figure presents the data from 16 two-year-olds. As can be seen, two-year-olds appropriately predict continued searching for Finds-Nothing and Finds-Substitute story characters but not for Finds-Wanted. They appropriately predict happiness for Finds-Wanted but sadness for Finds-Nothing and Finds-Substitute stories.

I propose that very young children are simple desire psychologists rather than belief–desire psychologists. Certainly the ease with which young children solved these simple desire judgement tasks is suggestive. In contrast we and others have found that two-year-olds almost universally fail belief-reasoning tasks. It only remains, therefore, to show that the same two-year-olds succeed at desire-reasoning tasks on the one hand but also fail comparable belief–desire reasoning tasks on the other. We (Wellman and Woolley, 1990) have demonstrated this in an experimental study with tasks like those described in figure 2.5. Rather than describe that study, however, I present some similar data from young children's everyday speech.

Natural Language Evidence

If it is true that a simple desire psychology precedes a belief–desire psychology, then there should be a young age when children talk about desires cogently but never talk sensibly, or never at all, about beliefs.

In the study I preview here Karen Bartsch and I are examining children's first use of desire terms – such as *want* and *wish* – and belief terms – such as *think* and *know* – in their everyday speech, studied longitudinally from approximately two to five years. To do this we used the longitudinal English corpora of utterances from ten children in the CHILDES database (MacWhinney and Snow, 1985). This datbase includes, for example, the utterances of Adam, Eve and Sarah from Roger Brown's research programme, and similar data from seven other children. Almost 400,000 child utterances are included in the transcripts we searched. We searched these transcripts for children's use of the desire terms, *want*, *wish*, *hope*, *afraid*, *care* (about); and for the belief terms, *think*, *know*, *expect*, *wonder*, *believe*, *understand*. Other studies (Bretherton and Beeghly, 1982; Shatz, Wellman and Silber, 1983) have shown that these are the earliest appearing terms of the sort we sought. Approximately three per cent of children's utterances included one or the other of these terms, so the sample of utterances of interest to us includes more than 10,000 child utterances.

The question we wish to address is not when children begin to use these terms but when children begin to use them for psychological reference. By psychological reference we mean use of the terms explicitly to talk of their own or others' internal, subjective states as distinct from external, objective aspects of behaviour and events. More specifically, when do children start to use these terms to talk about desire and to talk about belief? Of course, it is possible to use the terms in other fashions, to use them conversationally, as we will call it (see also Shatz et al., 1983), without really referring to beliefs and desires. Children can and do say things like 'know what?' using *know*, but without really referring to knowing. Or 'I want the toy', simply to request a toy, meaning no more than 'give me that toy', and not really referring to desire. But it is also possible to identify instances where the terms are used to genuinely refer to beliefs and desires amidst children's conversational uses.

In this study we carefully coded children's utterances into three large categories: (1) instances of genuine psychological reference, (2) conversational uses, and (3) uninterpretable uses. Instances of psychological reference were further coded as to whether they referred to belief and/or desire, and then as to their specific sense, for example, think, know, want, wish. To make a long story short, I will focus only on children's use of the term *want* as a term for desire, and *think* and *know* as terms for belief. This is a reasonable simplification because *want* was the term earliest used to refer to desires, and *think* and *know* were the terms earliest used to refer to beliefs. Further, in this sample of very young children's speech, 97 per cent of all expressions of desire used the term *want*, and 94 per cent of all expressions of belief used *think* or *know*.

Box 2.2 shows some sample utterances that were coded as genuine expressions of desire and of belief. In most instances, such as those shown here, identifying genuine reference to belief and desire was not too difficult given the precise coding procedures we used and given examination of each child's extended discourse. Inter-coder reliabilities were very near 90 per cent.

What are the findings? Figure 2.6 shows the occurrence of genuine psychological uses of *want*, *think* and *know* summed across all children. Note that genuine reference to a character's desire via the term *want* begins quite early and is well established even before the second birthday. Reference

Box 2.2

	Sample Utterances Expressing Desire
Dad:	What happened to your foot?
Child:	It hurt.
Dad:	Broken? Or cut?
Child:	*You want to see?*
Dad:	No, I'll see it later.
Child:	I want to show you.
Child:	Pants on. Jacket. Shopping
Mother:	We'll stay home and play.
Child:	*I want to shop . . . shopping.*
Child:	Santa Claus. Santa Claus is in town . . . street.
Mother:	It's not time yet.
Child:	*Wanna go see him.*
Child:	*Fraser (someone else) wants more coffee.*
Mother:	I'll put coffee on the fire, thank you.
Father:	What is it you're looking for?
Child:	*Want find top.* No top.
Child:	Here (offers object). That OK? *You want that?*
	Sample Utterances Expressing Belief
Child:	*Which one you think could fit?* This one.
Mother:	I think that one's a little large. Oh. I guess it does fit.
Child:	You were wrong
Child:	Some people don't like hawks. *They think they have . . . they are slimy.*
Mother:	What do you think?
Child:	I think they are good animals
Mother:	Where is that thing?
Child:	*I think it disappeared. You think it disappeared?*
Mother:	Yeah, I think it disappeared.
Child:	*I didn't know you had this.* Where did you get it?
Child:	Can I put my head in the mail box . . . *So the mailman can know where I are.*

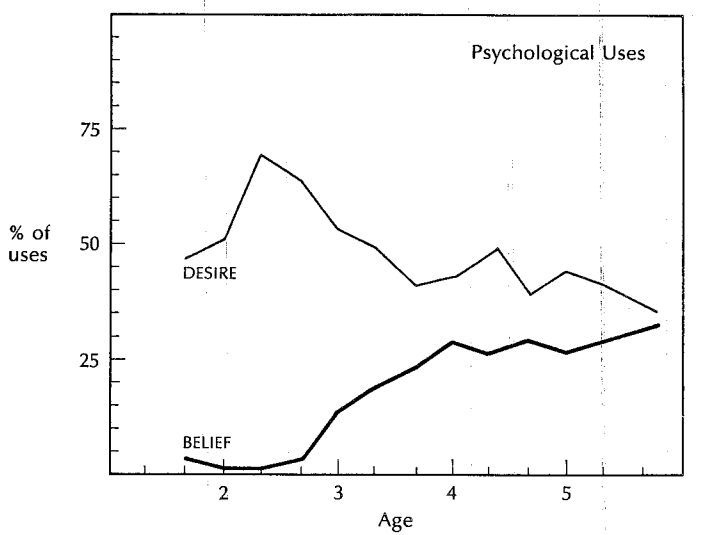

Figure 2.6 *Natural language occurrences of verbs of desire* (want) *and belief* (think *and* know) *used for psychological reference, as a percentage of all uses of these verbs.*

to belief via the terms *think* and *know* begins much later, just before the third birthday. More impressively, each individual child's data substantiate this general pattern. In every case (1) reference to desire precedes reference to belief, (2) reference to desire is already evident by two years, but (3) reference to belief is evident at just about three years.

In addition to these general codings, we were observant for more exacting and precise uses. For example, sometimes when children mention a belief or a desire they explicitly note that there is a difference between their own desire, say, and someone else's (or their own belief and someone else's). For example, 'Do you want me to look both ways? I don't wanna look both ways?' Or, as shown in box 2.2, 'They think they are slimy. I think they are good animals.' Such explicit self–other distinctions are rare in percentage terms but given our large data set were modestly frequent in absolute terms.

Figure 2.7 presents these data. Again, children refer to desires long before they refer to beliefs. That is, two-year-olds can clearly distinguish their own internal states from others' – they do so for desires. But only at about three years, again, do they explicitly distinguish their own beliefs from those of

Figure 2.7 *Natural language occurrences of verbs of desire (want) and belief (think and know) used to distinguish the mental states of self and other, as a percentage of all uses of these verbs.*

others. This corroborates, with more exacting data, the general picture: Namely, evidence of a rich understanding of simple desires *before* a genuine understanding of belief that emerges at about three years.

Conclusions

My discussion thus far requires some qualifications and conclusions. First, I am not saying that three-year-olds' belief–desire psychology and concomitant theory of mind is identical to our everyday adult one. It is not. Still, it is recognizably familiar, it is a mentalistic belief–desire psychology. At least one further major transition is required after three years (see Wellman, 1990 chapter 9). I will not discuss that transition here because I have concentrated instead on the prior transition to a theory of mind in the first place. About that transition I have a simple story to tell, namely that acquisition of a theory of mind takes place in normal humans at just about three years of age. It is apparent in three-year-olds' mentalism; it is preceded by and built

upon an earlier simple desire psychology. However, when I say that an understanding of belief and hence a belief–desire psychology and hence a concomitant theory of mind are acquired at age three I do not wish to be taken too literally. I do not believe that the conceptual developments I am charting are tied to chronological age at all, in any very precise way, I simply use these ages as convenient markers for talking about a sequence of early developments, namely the transition from an understanding of simple desires to an understanding of beliefs.

How does this transition take place? I do not know, but I can advance a plausible story of how it might work. In this analysis belief–desire psychology represents a theory change sponsored by and derived from simple desire psychology. Simple desire psychology provides the young child with significant explanatory resources, allowing the child to predict and understand a variety of actions and emotional reactions as stemming from the actor's internal desire states. However, a revision of desire psychology would be necessitated by predictive and explanatory failures of that reasoning scheme, failures which engender a construct of belief. For example, two characters with equal desires often engage in different actions. They do so because they have different beliefs. Same desires leading to different actions is a commonplace occurrence. It is easily accounted for in belief–desire psychology but is a theoretical anomaly for desire psychology. Similarly, actors often do things that *thwart* their own desires. Recall Jane searching for her kitty. That she should look for it under the piano when it is under the chair, is an anomaly for simple desire psychology, albeit easily accounted for by Jane's beliefs. Note, that thinking about actors in internal-state terms at all, that is with respect to their internal desires, makes it possible for the child to confront such theoretical anomalies in the first place. A behaviourist, for example, would not find these examples perplexing, he or she could easily account for them via histories of conditioning. It is desire psychology that generates *these* anomalies. Such anomalies once generated require addition of a very different sort of internal state construct to one's theoretical arsenal, specifically a conception of cognitive states of representation.

As this hypothetical scenario reveals, I believe that children's understanding of belief–desire psychology functions as a scientific theory functions in several important regards. Hence, I endorse the description of it as a theory of mind. One quality of scientific theories worth emphasizing is that they organize coherently a mass of potentially overwhelming, anomalous, even chaotic observations, experiences, data. Scientific theories do not appear out of thin air, because we have no facts, no observations, no first-hand data on which to rely. They appear just because we have so much data and observation to understand. Theories organize an otherwise immeasurable amount of data.

I wish to be clear about this characterization of theories because some people object to the phrase 'theory of mind' to describe three-year-olds' understanding (see chapter 1). Their objection seems to be that children's knowledge of beliefs and desire cannot be *theoretical* because children do not postulate beliefs and desires from the armchair, out of nothing, 'theoretically' as it were. Instead, so this objection goes, children *experience* their own beliefs, desires, emotions, perceptions (see Johnson, 1988; Harris, and also Butterworth, this volume). These are not theoretical entities for children, it is claimed, but first-hand experiences extended to make sense of others. I agree that children's database in this realm is large and immediate. Even infants accrue an extensive history of first-hand experience of their own cognition, motivations and internal states; they, further, amass innumerable experiences of interaction with and observation of others who, of course, display and express emotions, cognitions, and desires (see Wellman, 1990, chapter 8). However, the wealth of this experience does not mean that children's understanding is not theory-like. On the contrary, because there is so much experience and information available to children in so many punctate and co-mingled episodes and events, they must develop a coherent conceptual framework for organizing and distilling these experiences. This is the function of scientific theories, this is also the function of belief–desire psychology; this is the sort of theoretical understanding that children have achieved, substantially, by age three. Belief–desire psychology is a theory in this sense; it is an everyday theory forged from innumerable data, rather than constructed in absence of data.

Finally, how might my description of development of a theory of mind in humans shed light on the topics addressed by others in this volume? If I am right then the primitive condition for humans is neither behaviourism nor a theory of mind; it is instead a nonmentalistic but nonbehaviouristic, internal state theory. This might well prove true more generally. We may find that apes, for example, and autistic persons as another example, understand conspecifics as internal state organisms rather than cognitive ones. If so, we need to know this. We need to achieve a positive characterization of such subjects' understanding of others rather than simply asking if it is identical to our own and rather than simply showing that it is not a behaviourism. There is an important middle ground between these two extremes. For two-year-olds it is an impressive achievement to understand others in simple desire terms even if that does not constitute an understanding of mind. It would be a notable achievement for many other organisms as well.

3

Developing Understanding of Desire and Intention

JANET W. ASTINGTON AND ALISON GOPNIK

There is some evidence that two- and three-year-olds and autistic children understand others' desires and intentions (Baron-Cohen, chapter 16; Premack and Dasser, chapter 17; Wellman, chapter 2), while other evidence points to the serious limitations in such children's understanding of others' beliefs. For example, three-year-olds and autistic children cannot make predictions based on false beliefs (Leslie, chapter 5; Perner, Leekam and Wimmer, 1987), and three-year-olds cannot recognize change in their own beliefs (Gopnik and Astington, 1988), understand the distinction between appearance and reality (Flavell, 1986), nor identify the sources of their beliefs (Gopnik and Graf, 1988). Children normally succeed on all these tasks by five years of age. A number of authors have suggested, in slightly different ways, that three-year-olds and autistic children fail because they do not appreciate the representational nature of beliefs and their causal relationship to the world (Flavell, 1988; Forguson and Gopnik, 1988; Gopnik, 1990; Leslie, 1988a; Perner, 1988a).

Beliefs, desires, and intentions are all intentional mental states (Whiten and Perner, chapter 1) and so we might expect that children would come to understand them at the same point in time. These states all involve relationships to representations of reality, not to reality itself. However, this relation is different for beliefs and for desires and intentions. The difference is in terms of the direction of fit and the direction of causation between the representation and reality (Searle, 1983). For beliefs, the mind has to fit the world, that is, beliefs are true if the representation matches reality. If the representation does not match reality the belief is changed, and thus events in the world cause our beliefs. On the other hand, for desires and intentions, the direction of fit is from the world to the mind, for example,

The research reported in this chapter was supported by NSERC, Canada.