Mind & Language
Vol. 7 Numbers 1 and 2 Spring/Summer 1992

ISSN 0268-1064

Article
Why the Child's Theory of Mind Really Is a Theory

ALISON GOPNIK AND HENRY M. WELLMAN

How do children (and indeed adults) understand the mind? In this paper we contrast two accounts. One is the view that the child's early understanding of mind is an implicit theory analogous to scientific theories, and changes in that understanding may be understood as theory changes. The second is the view that the child need not really understand the mind, in the sense of having some set of beliefs about it. She bypasses conceptual understanding by operating a working model of the mind and reading its output. Fortunately, the child has such a model easily available, as all humans do, namely her own mind. The child's task is to learn how to apply this model to predict and explain others' mental states and actions. This is accomplished by running simulations on her working model, that is observing the output of her own mind, given certain inputs, and then applying the results to others.

The first position has a certain prominence; research on children's understanding of mind has come to be called 'children's theory of mind'. This position is linked to certain philosophers of mind such as Churchland (1984) and Stich (1983) who characterize ordinary understanding of mind, our mentalistic folk psychology, as a theory. It is also part of a recent tendency to describe cognitive development as analogous to theory change in science (Carey, 1985, 1988; Karmiloff-Smith & Inhelder, 1975; Wellman & Gelman, 1988; Keil, 1989; Gopnik, 1984, 1988). The second position, in

A. G. was supported by NSF grant BNS-8919916 during the preparation of this paper. H. W. was supported by a fellowship from the Center for Advanced Study in the Behavioral Sciences and grant HD-22149 from NICHD. We are grateful to Andrew Meltzoff, Clark Glymour, Chuck Kalish, and Doug Medin, who commented on an earlier version of this paper. Address for correspondence: Department of Psychology, University of California, Berkeley, California 94720. Email: gopnik@ucbcmsa.bitnet.

a somewhat different form, has a venerable philosophical tradition, going back to Descartes. This is the tradition of emphasizing the special importance of the first-person case in understanding the mind. More recently Gordon and Goldman have advocated a 'simulation theory' of mind (ST), and this position has been taken up in the developmental literature by Harris (1991, this issue) and Johnson (1988).

We do not believe that this is a dispute that can be settled on conceptual or *a priori* grounds. Rather it is a contest between two empirically testable hypotheses about the nature of 'folk psychology'. We believe that the child's understanding of mind is helpfully construed as a theory, and that changes in understanding may be thought of as theory changes. But we believe this because such an account provides the best explanation for the currently available developmental evidence.

In spite of the prominence of the 'theory theory' (TT), the exact nature of such folk psychological theories has rarely been spelt out in much detail, and in fact, this is often raised as an objection to this view. What exactly are the theoretical entities and laws that are involved in this theory? How is it constructed from the available evidence? We will first attempt to provide some of this detailed exposition. When the full story is told, we believe, a theory theory of early developments is compelling indeed. Second, we will argue that a contrasting simulation account fails to fit the data in key places and fails more generally to provide as comprehensive a view of development.

1. The Theory Theory

The question of what distinguishes a theory from other types of conceptual schemas is, of course, an enormous and difficult one. Nevertheless, it seems to us that there are characteristic features of both theories and theory change that can be outlined in very broad and simplified terms.

Theoretical constructs are abstract entities postulated, or recruited from elsewhere, to provide a separate causal-explanatory level of analysis that accounts for evidential phenomena. Gravity is not itself two bodies moving in relation to one another, it is postulated to explain such phenomena. Such theoretical constructs are typically phrased in a vocabulary that is quite different from the evidential vocabulary. For example, Kepler's theory of the planets includes ideas about elliptical orbits that are notoriously not visible when we look at the stars' motions in the sky. Theories in biology postulate unseen entities, like viruses and bacteria, with distinctive properties some of which are implicated in transmission of disease. Theoretical constructs need not be definitively unobservable, but they must be appeals to a set of entities removed from, and underlying, the evidential phenomena themselves. They are designed to explain (not merely type and generalize) those empirical phenomena. So, one characteristic of theories

is their abstractness. They postulate entities and analyses that explain the data but are not simply restatements of the data.[1]

Theoretical constructs do not work independently, they work together in systems characterized by laws or structure. A second characteristic of theories is their coherence. The theoretical entities and terms postulated by a theory are closely, 'lawfully', interrelated with one another.

The coherence and abstractness of theories together give them a characteristic explanatory force.[2] These features of theories also give them a very characteristic sort of predictiveness. To put it crudely, we can map a bit of evidence on to one part of the theory, grind through the intratheoretic relations, come out at a very different place in the theory and then map back from that part of the theory to some new piece of evidence. In this way, the set of abstract entities encompass a wide range of events, events that might not even seem comparable at the evidential level of description. A theory not only makes predictions, it makes predictions about a wide variety of evidence, including evidence that played no role in the theory's initial construction. Kepler's account allows one to predict the behavior of new celestial objects, moons for example, which were quite unknown at the time the theory was formulated. Theories in biology allow us to predict that antibiotics will inhibit many bacterial infections, including some, like scarlet fever, that present none of the symptoms of an infected wound, or some, like Legionnaire's disease, that were unknown when the theory was formulated. They also allow us to predict that such drugs will be useless against viral infections, even when the symptoms of the viral infection are identical to those of a bacterial one.

Some of these predictions will be correct, they will accurately predict future events described at the evidential level, and will do so in ways that no mere empirical generalization could capture. Others will be incorrect.

---

[1] It is important to be as clear as possible about the way in which we take theoretical constructs to be abstract, unobservable, and postulated. We mean abstract in the sense of 'thought of apart from' observable particularities, we do not mean abstruse or merely ideal. By unobservable we mean not obviously a part of the evidential phenomena to-be-explained; not that theoretical entities are necessarily incapable of being observed in any fashion whatsoever. Thus, we could postulate that genes control inherited features such as eye color and height, in order to provide a theoretical account, and still fully expect that genes are observable in some fashion. It is simply that genes are not directly evident in, observable in, the phenomena of eye color and height themselves. Similarly, postulated does not mean conjured out of thin air, it means recruited for explanatory purposes from outside the evidential phenomena themselves. Thus (natural) selection can be postulated to account for the origin of species but at the same time selection can be fully concrete and observable, in the realm of human animal breeding for example. It is the recruitment of selection to account for natural speciation that is postulational, selection itself is not a mere postulated entity.

[2] On the theory theory, the ages of development are not crucial. In fact we would expect to find, as indeed we do, wide variation in the ages at which successive theories develop. We would expect to find similar sequences of development, however. We will use ages as a rough way of referring to successive theories.

Since theories go beyond the evidence, and since theories are never completely right, some of their predictions will be falsified. In still other cases the theory will make no prediction at all. In fact, the theory may in some circumstances have less predictive power than a large set of empirical observations. This is because explanatory depth and force do not simply equate with predictive accuracy. We can make predictions about events without explaining them: Kepler's theory still leaves many of Tycho Brahe's observations unexplained. The differences in cases of theoretical prediction are two-fold. First, a few theoretical entities and laws can lead to a wide variety of unexpected predictions. Second, in the case of a theory, prediction is intimately tied to explanation.

An additional characteristic of theories, related to this central function of explanation, is that they produce interpretations of evidence, not simply descriptions of evidence and generalizations about it. Indeed theories influence which pieces of evidence we consider salient or important. In modern medicine, for example similar sets of symptoms do not necessarily yield the same disease diagnosis. An empirical typology of similar symptoms is overriden by deeper more theoretic biological explanations. The interpretive effects of theories may be stronger still, it is notoriously true that theoretical preconceptions may lead a scientist to dismiss some kinds of evidence as simply noise, or the result of methodological failures. Nor is this simply prejudice. On the contrary, deciding which evidence to ignore is crucial to the effective conduct of a scientific research program.

All these characteristics of theories ought also to apply to children's understanding of mind, if such understandings are theories of mind. That is, such theories should involve appeal to abstract unobservable entities, with coherent relations among them. Theories should invoke characteristic explanations phrased in terms of these abstract entities and laws. They should also lead to characteristic patterns of predictions, including extensions to new types of evidence and false predictions, not just to more empirically accurate prediction. Finally, theories should lead to distinctive interpretations of evidence, a child with one theory should interpret even fundamental facts and experiences differently than a child with a different theory.

So far we have been talking mostly about the static features of theories, the features that might distinguish them from other cognitive structures such as typologies or schemas. But a most important thing about theories is their defeasibility. Theories are open to defeat via evidence and because of this theories change. In fact, a tenet of modern epistemology is that any aspect of a theory, even the most central ones, may change. The dynamic features of theories, the processes involved in theory formation and change, are equally characteristic and perhaps even more important from a developmental point of view.

While any very precise specification, any algorithm, for theory change may elude us, there are certainly substantive things to be said about how it typically takes place. There are characteristic intermediate processes

involved in the transition from one theory to another. One particularly critical factor is the accumulation of counter-evidence to the theory. The initial reaction, as it were, of a theory to counter-evidence may be a kind of denial. The interpretive mechanisms of the theory may treat the counter-evidence as noise, mess, not worth attending to. At a slightly later stage the theory may develop *ad hoc* auxiliary hypotheses designed to account specifically for such counter-evidence. Auxiliary hypotheses may also be helpful because they phrase the counter-evidence in the accepted vocabulary of the earlier theory. Such auxiliary hypotheses, however, often appear, over time, to undermine the coherence that is one of a theory's strengths. The theory gets ugly and messy instead of being beautiful and simple.

A final step requires the availability or formulation of some alternative model to the original theory. A theory may limp along for some time under the weight of its auxiliary hypotheses if no alternative way of making progress is available. But the fertility of the alternative idea itself may not be recognized immediately. Initially it may only be applied to the problematic cases. We may see only later on that the new idea also provides an explanation for the evidence that was explained by the earlier theory.

The development of the heliocentric theory of the planets provides some good examples of these processes. Auxiliary hypotheses involving more and more complex arrangements of epicycles were initially invoked to deal with counter-evidence. Later heliocentrism was introduced by Copernicus. It is worth noting though that Copernicus' theory fails to apply the central heliocentric idea very widely. In many respects Copernicus' account is more like the Ptolemaic ones, than, say, Tycho Brahe's account. It includes epicycles, for example. Brahe's account acknowledges many of the flaws of the Ptolemaic ones, and uses the idea of heliocentrism to deal with them (other planets revolve around the sun which revolves around the earth). But Brahe fails to accept the central idea that the earth itself goes round the sun. Only with Kepler is there a really coherent heliocentric account that deals both with the anomalies and with the earlier data itself.

We propose that these same dynamic features should be apparent in children's transition from one theory to a later one, and specifically from one view of the mind to another. Children should ignore certain kinds of counter-evidence initially, then account for them by auxiliary hypotheses, then use the new theoretical idea in limited contexts, and only finally reorganize their knowledge so that new theoretical entities play a central role.

2. *The Child's Theories of Mind*

We propose that there is a change from one mentalistic psychological theory to another somewhere between $2\frac{1}{2}$ and around 4.[2]. The change is not a simple all-or-none one, but rather involves a more gradual transition from one view of the mind to another. Indeed this change manifests the

telltale intermediate processes that are characteristic of theory change. Two-year-olds have an early theory that is incorrect in that it does not posit the existence of mental representational states, prototypically beliefs. In 3-year-olds there is an intermediate phase where children demonstrate an understanding of the existence of representational states, at times, but only as auxiliary hypotheses. That is children in this phase can acknowledge that representational states of mind exist, if forced to do so in certain ways, but this realization is peripheral to their central explanatory theory. In a third phase, beginning around 4, children reorganize their central explanatory theory, it becomes properly a belief-desire psychology. Children begin to realize that what the actor thinks—his or her representation of the world rather than the world itself—inevitably determines actions.

2.1 The 2-year-old Theory

The 2-year-old is clearly a mentalist and not a behaviorist. Indeed, it seems unlikely to us that there is ever a time when normal children are behaviorists. Even in infancy, children seem to have some notions, however vague, of internal states as evidenced in early primary intersubjectivity (Trevarthen & Hubley, 1978) and imitation (Meltzoff & Moore, 1977; Meltzoff & Gopnik, in press) and later, more clearly, in social referencing and joint attention behaviors (*e.g.* Wellman, in press). It seems plausible that mentalism is the starting state of psychological knowledge. But such primary mentalism, whenever it first appears, does not include all the sorts of mental states that we as adults recognize. More specifically, even at two years psychological knowledge seems to be structured largely in terms of two types of internal states, desires, on the one hand and perceptions, on the other. However, this knowledge excludes any understanding of representation.

Desire and perception alone provide examples of the two basic categories of explanatory entities in folk psychology—the two types of theoretical constructs that Searle calls 'world-to-mind' and 'mind-to-world' states (Searle, 1983). An understanding of desire encompasses an early knowledge that what's in the mind can change what's in the world. An understanding of perception, on the other hand, encompasses an early knowledge that what's in the mind depends on what's in the world. Moreover, both desire and perception, as theoretical constructs, work to explain action but may also be divorced from any particular actions that an agent may perform.

Importantly, however, desire and perception can be, and at first are, understood in nonrepresentational terms. Desires at first are conceived simply as drives towards objects (Wellman & Woolley, 1990). Perceptions are at first understood simply as awareness of objects (Flavell, 1988). In neither case need the child conceive of a complex propositional or representational relationship between these mental states and the world. Instead, these very young children seem to treat desire and perception as fairly simple causal links between the mind and the world. Given that an

agent desires an object, the agent will act to obtain it. Given that an object is within a viewer's line of sight, the viewer will see it. These causal constructs are simple, but they have considerable predictive power. In particular, together they allow the first form of 'the practical syllogism': 'If an agent desires X, and sees it exists, he will do things to get it'. Even that form of the practical syllogism is a powerful inferential folk psychological law. It allows children to infer for example, that if John wants a cookie and sees one in the cookie jar, he will go there for it. If he doesn't want it, or doesn't see it, he won't.

2.2 The 3-year-old Theory

By three, children begin to show signs of a more elaborate mental ontology. Given the difficulties of testing children younger than three, the earliest emergence of this aspect of the theory is difficult to document. While 2-year-olds successes on desire and perception tasks are striking, their failures on other tasks are more difficult to interpret. However, natural language can provide us with one avenue for exploring these abilities. Before three, children make extensive and appropriate use of terms for desire and perception (Bretherton & Beeghly, 1982). More cognitive mental terms (think, know, remember, make-believe, dream) only begin to emerge at around the third birthday (Shatz, Wellman & Silber, 1983).

There is further evidence that at three children begin to have a more general notion of belief and also of such representational but 'not real' mental states as pretenses, dreams, and images (*e.g.* Wellman & Estes, 1986). When these concepts first appear, however, they have an interesting character, framed by the child's larger theory which is still a desire-perception theory. This manifests itself in two ways. First, understanding of belief appears to be initially modelled on a non-representational understanding, that is modelled on an earlier understanding of desire and especially perception. Second, even when the notion of belief, as a representation, appears it first plays little if any role in the child's explanations of behavior. In these respects the child's first conception of belief seems to be a conceptual construction based on reworking earlier theoretical constructs. Moreover, even the more advanced representational notion initially functions like an auxiliary hypothesis rather than a central theoretical construct.

To elaborate, 3-year-olds' first understanding of belief seems like their earlier understanding of perception in that it shares something of that construct's nonrepresentational character. Specifically, belief does not at first easily encompass a sense of misrepresentation. On this view, belief, like perception and desire, involves rather direct causal links between objects and believers. This view has variously been called a 'copy theory' (Wellman, 1990), a 'Gibsonian theory' (Astington & Gopnik, 1991a) a 'situation theory' (Perner, 1991), or a 'cognitive connection' (Flavell, 1988) theory of belief. The similar idea in all these accounts is that belief contents

directly reflect the world. The introduction of a notion of belief promises an important additional complexity to the child's theory of mind. Initially however, the notion seems to be quite strongly embedded in the nonrepresentational desire-perception framework of the earlier theory.

At times, however, at least as the fourth year progresses, 3-year-olds are able to recognize the existence of beliefs that clearly misrepresent. They can explain already completed, ineffective actions as indicating a false belief by the actor and can at times even acknowledge the presence of mistaken, wrong beliefs (*e.g.* Siegal & Beattie, 1991; Moses, 1990). However, these same children do not often construe actions as stemming from false beliefs. When predicting action they typically, consistently, resistantly act as if the actor's desire along with the objective facts determine action, ignoring a role for false belief in influencing action (*e.g.* Gopnik & Astington, 1988; Perner, Leekham & Wimmer, 1987; Wellman & Bartsch, 1988). Similarly, when asked the contents of person's belief, they consistently, resistantly cite the facts (*e.g.* Perner *et al.*, 1987; Moses & Flavell, 1990). In short, when predicting action and when diagnosing belief contents, 3-year-olds evidence largely a nonrepresentational desire-perception understanding.

What about 'non-real' mental states, such as pretences, dreams, and images? There is evidence that children actually have such fictional mental states as young as 18 months (*e.g.* Leslie, 1987). Evidence that they understand such states, however, is much less clear. By the third birthday, however, children have some conceptual knowledge of these aspects of mental life (*e.g.* Wellman & Estes, 1986; Harris, Brown, Marriot, Whittall & Harmer, 1991). Moreover, they may distinguish such imaginary or hypothetical states from the states of desire and perception. However, these states appear to play little role in children's explanation of ordinary behavior. More significantly, these states have little causal connection to objects (that, in fact, is what is distinctive about them). While children see desires as states that modify the world, and perceptions as states that are modified by the world, pretenses, images and dreams, on their view, bear no causal relation to the world at all. It is possible that postulating these states, which are representational but divorced from reality, also plays a role in the eventual development of the full representational theory.

In summary, mental representations exist for 3-year-olds, but only as a relatively isolated auxiliary hypothesis necessary to explain certain (to them) peripheral mental phenomena—the odd infrequent misrepresentation and explanatorily impotent fictional representations.

2.3 The 5-year-old View

By four or five, children, at least in our culture, have developed a quite different view of the mind, one that we have called a 'representational model of mind' (Forguson & Gopnik, 1988). On this view almost all psychological functioning is mediated by representations. Desires, percep-

tions, beliefs, pretences and images all involve the same fundamental structure, a structure sometimes described in terms of propositional attitudes and propositional contents. These mental states all involve representations of reality, rather than realities themselves. In philosophical terms, the child's view of the mind becomes fully 'intentional'. To use Dretske's terminology perceiving becomes perceiving that, and desiring becomes desiring that, we might even add, that believing becomes believing that (Dretske, 1981). This new view provides a kind of Copernican, or better Keplerian, revolution in the child's view of the mind. In addition to distinguishing different types of mental states with different relations to a real world of objects, the child sees that all mental life partakes of the same representational character. Many characteristics of all mental states, such as their diversity, and their tendency to change, can be explained by the properties of representations. This newly unified view not only provides new predictions, explanations and interpretations; it also provides a new view of the very evidence that was accounted for earlier by the desire-perception theory.

3. The Child's Theory as Theory

What evidence do we have for thinking that these understandings are theoretical in the sense that we have been outlining so far? The following: The child's understanding involves general constructs about the mind that go beyond the focal evidential phenomena. These constructs feature importantly in explanation. They allow children to make predictions about behavior in a wide variety of circumstances, including predictions about behavior they have never actually experienced and incorrect predictions. Finally, they lead to distinctive interpretations of evidence.

3.1 Explanations

Children's explanations of actions show a characteristic theory-like pattern. In open-ended explanation tasks (Bartsch & Wellman, 1989; Wellman & Banerjee, 1991) children are simply presented an action or reaction ('Jane is looking for her kitty under the piano') and asked to explain it ('Why is she doing that?'). There are many mental states that might be associated with such situations. Yet 3- and 4-year-old children's answers to such open-ended questions are organized around beliefs and desires just as adults' are ('she wants the kitty'; 'she thinks it's under the piano'). Moreover, there is a shift in explanatory type between two and five. Two-year-olds' explanations almost always mention desires, but not beliefs. Asked why the girl looks for her doll under the bed they will talk about the fact that she wants the doll, but not the fact that she believes the doll is there. Three-year-olds invoke beliefs and desires, and some threes and most 4- and 5-year-olds consistently refer to the representational character of these

states, explaining failure in terms of falsity. These same trends can be seen in the explanations children give in their spontaneous speech (Bartsch & Wellman, 1990).

3.2 Predictions

Consider the desire-perception theory. Even that early theory allows children to make a variety of predictions about actions and perceptions, both their own and others. For example, they should be able to predict that desires may differ, and that, given a desire, an actor will try to fulfil that desire. They should know that desires may not be fulfilled. They should predict that fulfilled desires will lead to happiness, while unfulfilled desires will lead to sadness (Wellman & Woolley, 1990). And there is evidence that, in fact, all these kinds of predictions are made by very young children (*e.g.* Wellman & Woolley, 1990; Yuill, 1984; Astington & Gopnik, 1991b). Similarly, a child with the desire-perception theory should be able to predict the perceptions of others in a wide variety of circumstances, including those in which the perceptions are different from their own. Such very early activities as shared attention and social referencing behaviors already indicate some capacity to understand the perception of others (Wellman, in press). Other aspects of this understanding quickly develop. By $2\frac{1}{2}$ these Level-1 understandings, as Flavell calls them, are firmly and reliably in place (Flavell, 1988). At this age children can reliably predict when an agent will or will not see (and hear and touch) an object (*e.g.* Flavell, Everett, Croft & Flavell, 1981). They can also predict how seeing an object will lead to later actions. However, they are unable to make predictions about representational aspects of perception, what Flavell calls Level-2 understanding. They fail to predict, for example, that an object that is clearly seen by both parties can look one way to one viewer and another way to another.

These predictions may seem so transparent to adults that we think of them not as predictions at all but simply as empirical facts. A little reflection, however, should make us realize that the notion of desire or perception used by these very young children is theoretically broad and powerful. Children can use the notion of desire appropriately and make the correct predictions when the desired objects are objects, or events, or states of affairs. They can attribute desires to themselves and others even when they do not act to fulfill the desires and when the desires are not in fact fulfilled. Similarly, children seem to make accurate predictions about perception across a wide range of events, involving factors as different as screens, blindfolds, and visual angles, and do so across different perceptual modalities. Again, they may do so even when the perceptions do not lead to any immediate observable actions. Moreover, given novel and unfamiliar information about an agent's desires and perceptions, children will make quite accurate predictions about the agent's actions.

More significantly, however, these children also make incorrect predic-

tions in cases where the desire-perception theory breaks down. Both desires and perceptions, on the 2-year-old view, involve simple non-representational causal links between the world and the mind. Even the early non-representational notions of belief have this quality. This theory cannot handle cases of misrepresentation. Presented with such cases it makes the wrong predictions. The theory also cannot handle other problems that require an understanding of the complexity of the representational relations between mind and world. For example, the theory breaks down when one must consider the fact that the same belief may come from different sources, or that there may be different degrees of certainty of beliefs.

The most well-known instance of such an incorrect prediction is, of course, the false-belief error in 3-year-olds (Wimmer & Perner, 1983; Perner *et al.*, 1987). The focus on false belief tasks may, however, be somewhat unfortunate since it has promoted a mind-set in which any ability to perform 'correctly' on a false-belief task is taken as evidence that the child has a representational theory of the mind. As we will see, there are cases in which 3-year-olds indicate some understanding of false belief. However, to begin with it is worth pointing out the much greater ubiquity and generality of the incorrect false-belief predictions. Three-year-olds make erroneous predictions, not only in the 'classic' tasks, but also in many other cases involving beliefs about location, identity, number and properties. They make incorrect predictions for 'real' others, for puppets, for children, and for hypothetical story characters. Incorrect predictions are made when the question is phrased in terms of what the other thinks, what the other will say and what the other will do, and across a wide range of syntactic frames. They are made by North American (Gopnik & Astington, 1988), British (Perner *et al.*, 1987), and Austrian (Wimmer & Perner, 1983) children, and recently by Baka children of the Cameroons (Avis & Harris, 1991).

Moreover, and more significantly from the point of view of the theory theory, these incorrect belief predictions are mirrored in 3-year-olds' performance on a wide range of other tasks. A brief inventory would include (a) appearance-reality tasks, which themselves have proved robust across many variations of culture, question and material (Flavell, Green & Flavell, 1986), (b) questions about the sources of belief (Gopnik & Graf, 1988) and the understanding of subjective probability (Moore, Pure & Furrow, 1990), and (c) the understanding of pictorial representational systems (Zaitchek, 1990). In some of these tasks the desire-perception theory makes incorrect predictions, and children consistently give the same wrong answer. In others, it makes no predictions at all and the children respond at random. On any information-processing account these tasks would require quite different kinds of competences. Moreover, the standard methodology of these studies has included control tasks, involving similar or identical information-processing demands, which children seem entirely capable of answering. Nor do any dimensions of familiarity, at least in any simple

terms, seem to underlie the difference between tasks at which children succeed and fail.

3.3 Interpretations

In these cases children are clearly using belief and desire to make predictions—one of the central functions of theoretical constructs. In addition to the explanatory and predictive effects, children also show strong interpretive effects. Suppose we present the child with counter-evidence to the theory? If the child is simply reporting her empirical experience we might expect that she will report that evidence correctly. In fact, however, children consistently misreport and misinterpret evidence when it conflicts with their theoretical preconceptions. Flavell and his colleagues have some provocative but simple demonstrations of evidential misinterpretation (Flavell, Flavell, Green & Moses, 1990). A child sees a blue cup, agrees that it is blue and not white, and sees the cup hidden behind a screen. At this point another adult comes into the room, and she says 'I cannot see the cup. Hmm, I think it is white'. Then the child is asked what color he thinks the cup is and what color the adult thinks it is. To be correct the child need only report the adult's actual words, but 3-year-olds err by attributing to the character a true belief. Even if corrected, 'well actually she really thinks it's white', 3-year olds continue to insist the adult has a factually correct belief: 'She thinks it's blue'. Moreover, as we will see, three-year-old children consistently misreport their own immediately past mental states.

3.4 Transitional Phenomena

In developmental psychology we are often better at describing the states at two points in development than at describing changes from one state to another. Nevertheless, 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. This is rather bad luck for developmentalists since this period has been the focus of much of our investigation. But it also means that we may have some intriguing evidence about the mechanisms that lead from one theory to another.

We have already seen in our discussion of interpretation how children with the earlier theory begin by simply denying the existence of the counter-evidence. Johnny and I really did think and act as if there were pencils in the box when we first saw it. We have also seen that at around three children develop a first non-representational account of belief, which extends their original desire-perception psychology. We can also ask where the first signs of an understanding of misrepresentation, the centerpiece of the 5-year-old theory, begin to appear. Recall that we suggested, in the scientific case, that in a transitional period the crucial idea of the new

theory may appear as an auxiliary hypothesis couched in the vocabulary of the original theory, or be used in order to deal with particularly salient types of counter-evidence, but may not be widely applied. There is evidence for both these phenomena in the period from three to four. Children seem to us to initially develop the idea of misrepresentation in familiar contexts like those of desire and perception, without extending the idea more generally. They also initially apply the idea only when they are forced to by counter-evidence.

There is evidence that by placing the misrepresentation questions in the context of the earlier theory we can begin to see (or perhaps, in fact, induce) glimmerings of the later theory. Desire and perception may be construed either non-representationally, or representationally. In fact, in the adult theory, desire and perception are as representational as belief. What we want and see (by and large) is not the thing itself but the thing as represented. Understanding some aspects of desire and perception requires this sort of representational understanding. When we are satiated with something we no longer desire it, but the object itself has not changed. When different types of people have different tastes or values, their desires differ but the objects of desire remain the same (Flavell *et al.*, 1990). There is evidence that these representational aspects of desire are understood earlier than equivalently representational aspects of belief (Gopnik & Slaughter, 1991). However, 3-year-old children still do not perform as well on these tasks as they do on simple nonrepresentational desire tasks. Similarly, while non-representational aspects of perception are understood by 2½, representational ones, what Flavell calls level-2 perspective-taking, are only understood later (Flavell *et al.*, 1981; Flavell *et al.*, 1986; Masangkay, McCluskey, McIntyre, Sims-Knight, Vaughn & Flavell, 1974). However, there is evidence that these aspects of perception are understood before corresponding aspects of belief. Both in Flavell's earlier studies and in a recent study we conducted (Gopnik & Slaughter, 1992), children were better at misrepresentation tasks involving perception than they were at similar appearance-reality and false-belief tasks.

We have suggested that for the 2-year-old the central theoretical constructs are non-representational desires and perceptions while for the 5-year-old they are representational beliefs. Three-year-old precursors seem to include both non-representational accounts of belief and representational accounts of desire and perception. This is reminiscent of the way that Copernicus and Tycho Brahe mix epicycles and heliocentrism.

There is also evidence that early signs of an understanding of misrepresentation may come when children are forced to consider counter-evidence to their theory. In particular, Bartsch and Wellman (1989) found, as others had, that 3-year-old children continued to make incorrect false-belief predictions even given counter-evidence. However, if children were asked to explain the counter-evidence, at least some of them began to talk about misrepresentation as a way of doing so. Making the counter-evidence particularly salient seemed to help to induce the application of the

theory in this transitional age group. Similarly, in a recent study, Mitchell and Lacohee (in press) found that children in a representational change task who selected an explicit physical token of their earlier belief (a picture of what they thought was in the box) were better able to avoid later misrepresentation of that belief. That is, these children seemed to recognize the contradiction between the action they had just performed (picking a picture of candies) which was well within the scope of their memory, and their theoretical prediction about their past belief. Some evidence from natural language may also be relevant. Before age three (or slightly earlier) we simply do not find genuine references to belief. At about three, however, we begin to see such references, and also to see beginnings of contrastive uses of belief terms (Bartsch & Wellman, 1990). These uses may occur in contexts in which some particularly salient piece of counter-evidence to the earlier theory takes place. During the following year, however, the use of these terms increases drastically.

In short, children seem to first understand both belief and representation as small extensions of the original non-representational desire-perception theory, essentially as auxiliary hypotheses. This stage appears to be an intermediate one between a fully non-representational and a fully representational theory of mental states.

Do 3-year-olds really understand false belief then? Did Copernicus really understand planetary movement? The answer in both cases is that the question is a bad one. One of the strengths of the theory theory is that it makes such questions otiose. 'Understanding' false belief, or developing an idea of representation, involves the development of a coherent, widely applicable theory. It may be possible to have some elements of that theory, or to apply them in some cases, without operating with the full predictive power of the theory, particularly in a transitional state.

We argue therefore, that the transition from 2½ to 5 shows all the signs of being a theory change. While initially the theory protects itself from counter-evidence, the force of such counter-evidence eventually begins to push the theory in the direction of change. The first signs of the theory shift may emerge when counter-evidence is made particularly salient. Moreover, the theory initially deals with such counter-evidence by making relatively small adjustments to concepts that are already well-entrenched, such as desire and perception. Finally, by 4 or 5 the new theory has more completely taken over from the old. The predictions are widely and readily applicable to a range of cases.

4. Simulation Theory

In the theory theory, to predict someone's behavior we have recourse to theoretical constructs such as beliefs and desires. Explaining someone's behavior involves more than empirical generalization (X has always done this in similar situations in the past). It involves appeal to constructs at a

very different level of vocabulary—X wants Y and believes Z. A distinction between a phenomenal description and a theoretical explanation is crucial.

On the simulation theory, however, the child's (and adult's) understanding of mind is more closely linked to the phenomenal than to the theoretical. Understanding states of mind involves empirically discovering the states or results of a model. Consider again an understanding of the planets. An appeal to theoretical notions such as heavenly bodies revolving around one another can be contrasted to use of a planetarium-model to predict the star's appearance. (Here we want to be careful to focus on a user of a planetarium who has no deeper understanding of its workings and not focus on the planetarium's creator, for example, who presumably understood something theoretical about planetary motion in order to build a successful device.) The user need only see, empirically, that the planetarium's behavior mimics the stars, then the user can make predictions by 'running' the planetarium rather than waiting for the actual events. And the user can achieve a sort of explanation, explanation-by-demonstration, as well. Let's say the user experiences a real eclipse for the first time, noting that in the middle of the day it very uncharacteristically gets dark, although there are no clouds in sight. 'Why?', he asks himself. Is this a breakdown in all his empirical generalizations about the system; is it to be expected again; what happened? By running the planetarium under appropriate conditions the user can 'see' the phenomenon again, see that it occurs regularly, in the model; see that it is a natural although infrequent empirical fact. If asked by someone else 'What was that (eclipse)?' or 'Why did that happen?' the user can explain-by-demonstrating: 'Look, it (the eclipse) was one of these (demonstrate the model's state). It happens when the other stars are like this.'

The simulation theory contends that our prediction and explanation of mental phenomena is like that of the planetarium user. The child (or adult) doesn't need and doesn't appeal to a theory of mind, a conceptual understanding of mental states, to predict behavior or understand others. Instead she simply runs a perfect working model of a mind, her own mind. By considering the output of her own mind she can predict the mental states and resultant behaviors of others. And to explain curious or unexpected actions she can run her model, find a suitable simulated demonstration of the phenomena, and then explain it as 'look, it's one of these'.

Consider, for example, the classic false-belief task. The child sees a candy box, finds out that it is full of pencils, and then is asked what another person will think is inside it. Simulation theorists contend that the child need not have anything like a theoretical construct of belief (or desire) to solve this task. She simply has access to her own first-hand mental system and uses that. When asked what the character 'thinks', she need not understand beliefs as something like a representational construct, she simply simulates the experience and reports her own specific resulting state—'Oh, I (she) think's there is candy in the box'. The earlier failure to solve this task, on this view, reflects a failure of simulation, rather than a

failure of knowledge. It is not that the younger child fails to understand beliefs as states of misrepresentation, as we described it earlier, it is just that the younger child makes an egocentric simulation, projecting her own current mental states onto the other, rather than adjusting the simulation to the other's particular condition.

The simulation view has a number of telling empirical consequences; we will focus on two. The first concerns the centrality of your own mind in any understanding of the minds of others. To answer questions about others, according to ST, you must conduct a simulation on a model, and that model is your own mind. On the simulation view, therefore, the outputs of your own mental system are particularly central to all discourse about the mind. Moreover, these outputs must be easily and transparently accessible. This must be true in order for the simulation account to work at all in the case of other people. A presupposition of the account is that it is possible to read off and report the output of your own mental states, and to use them in explanation, prediction and inference. Moreover, access to your own states requires no inference or interpretation, no conceptual intermediaries, no theorizing; you simply read them off. A consequence of this view is that one cannot erroneously misinterpret, or misconceive, one's own mental state. You could of course run a bad simulation, in the sense that you entered the wrong inputs. But given those inputs, the output must be accurate. It must accurately reflect what your mind would actually do in that situation, because it IS what your mind actually does in that situation.

On the theory view, in contrast, erroneous self-interpretations are not only possible, they are to be expected. One typical characteristic of theories, after all, is that they allow and often even force interpretation of the evidence. If the theoretical prediction and evidence are in conflict it is often the evidence rather than the theory which is reinterpreted. Equally, on the theory theory psychological constructs, such as beliefs and desires, are generically applicable to the self or to others. If you possess a faulty conception of some mental state, say belief, then you will incorrectly attribute that mental state to others, and you should make parallel incorrect attributions to yourself. In short on the Simulation Theory false interpretations of your own mental states should not occur. On the Theory Theory such false interpretations should occur whenever your theoretical constructs are faulty.

A second empirical consequence, related to this first one, concerns how development should proceed. For both TT and ST we can predict that there will be development: children should first be good at predicting/explaining 'easy' states and then later 'hard' ones. But the notions of easy and hard should differ dramatically between these two theories. For ST the critical difference should be between states that are difficult or easy to simulate. Presumably, the metric for such ease and difficulty must be intimately related to the similarity of the states to the child's own states. In this sense the simulation theory is in another long and honorable tradition, the

tradition of 'perspective-taking' views in development. Several of the simulation theorists in this Special Issue for example presume that young children's errors are 'egocentric'. That is, the child's early errors consist of not correctly adjusting their simulation to the other person's condition. Note that on this theory there is no reason to expect that different mental states should be easier or harder to attribute to others. Take beliefs and desires. Both beliefs and desires are equally available to the child as states of her own mind. At a young age we could predict that reading off one's own beliefs and desires should be equally easy, and attributing conflicting beliefs and desires to someone else should be equally difficult.

In contrast, for the theory theory the critical metric concerns states that are easy or difficult to conceive of. Earlier we described what we take to be a succession of changes in the child's conceptions of mental states, as the child develops and replaces a succession of theories. Especially important is a difference between an early non-representational understanding of mind and a later more representational understanding. Early on children have a relatively adequate understanding of non-representational desire-perception states. Later they develop an understanding of the representational state of belief, specifically, and a representational understanding of mind more generally (including a representational understanding of certain aspects of perception and desire). Theoretical conceptions of the sort we have described are equally applicable to the self and others. If a theory has formulated a particular theoretical construct, such as the concept of false representations, it should in principle be able to use this concept equally to explain the child's own behavior and the behavior of others. If the theory does not include this construct, it should not be so applicable to either the self or others. In short, for the theory theory it will not be so important whether the mental states to be reasoned about are those of self or other. What is important is the relevant conceptions of mental states that the child must bring to bear. Thus, we find different developmental predictions from the two theories.

We want to describe several empirical findings based on these two main issues that tell against the simulation account. (1) Three-year-old children make false attributions to themselves, that exactly parallel their false attributions to others. (2) Three-year-old children make correct non-egocentric attributions to themselves and others for some mental states. (3) Children refer to only some mental states in their explanations, and refer to different mental states at different stages of their development. (4) Children's understanding of other psychological phenomena changes in parallel with their understanding of false belief. Understanding these phenomena does not require simulation, but it does require a representational theory of the mind.

The first set of findings concern children's ability to understand and report their own mental states. For example, children not only fail to understand that other's beliefs can misrepresent; they also fail to understand that their own beliefs can. In our original experiment (Gopnik &

Astington, 1988) we used an analog of the 'false-belief' task. We presented children with a variety of deceptive objects, such as the candy box full of pencils, and allowed them to discover the true nature of the objects. We then asked the children the standard false-belief question, 'What will Nicky (another child) think is inside the box?'. But we also asked children about their own false beliefs about the box: 'When you first saw the box, before we opened it, what did you think was inside it?' The pattern of results for self and for other was very similar. 3-year-olds tend to say that Nicky will think what is true. But they also report that they themselves thought what was true, that they had originally thought there were pencils in the box. Children's ability to answer the false-belief question about their own belief was significantly correlated to their ability to answer the question about the others' belief, even with age controlled, a result recently replicated by Moore *et al.* (1990). Children who could not answer the question about the other, also could not answer it about themselves.

The children also received an additional control task. They saw a closed container (a toy house) with one object inside it, then the house was opened, the object was removed and a different object was placed inside. Children were asked 'When you first saw the house, before we opened it, what was inside it?'. This question had the same form as the belief question. However, it asked about the past physical state of the house rather than asking about a past mental state. Children were only included in the experiment if they answered this question correctly, and so demonstrated that they could understand that the question referred to the past and could remember the past state of affairs. Several different syntactic forms of the question were asked to further ensure that the problem was not a linguistic one. Recently, this experiment has been replicated, with additional controls, by Wimmer and Hartl (1991).

In more recent experiments we have investigated whether children could understand changes in mental states other than belief (Gopnik & Slaughter, 1991). A crucial comparison is to desires and perceptions. In three different tasks we presented children with situations in which their desires were satiated and so changed. For example, initially the child desired one of two short books. That one was read to him and the child said he now desired the other book. The test question was just like the one for past beliefs: 'When you first saw the books, before we read one, which one did you want?' In these tasks 3-year-old children were considerably better at reporting past now-changed desires than past now-changed beliefs. Similarly, we presented children with situations in which their perception was changed. Children saw an object on one side of a screen and they were then moved to the other side of the screen where they saw a different object. We asked 'When you first sat on the chair, before we moved over here, what did you see on the table?'. Children were completely able to report their past perceptions.

These experiments concern the child's report of their own mental states, beliefs, desires and perceptions. From a simulation point of view, why do

the child make errors when they are simply reading off their own mental states? And why do they make errors for one state but not the other? Perhaps the trouble is that the questions require not a report of current mental states, but a memory of past states. Two things need to be kept in mind in considering this objection. First, the span of time we are talking about is very brief, at the most one or two minutes and often much shorter. At least for adults such experiences are well within the immediate introspective span. If I were to report the output of my mental system in such a situation, I would report the change in my belief that comes with the new discovery, with all its attendant phenomenological vividness and detail. The very psychological experience of the change in belief depends on the fact that I continue to remember the previous belief. A simulation account must presuppose some ability to report immediately past states (after all any state will be past by the time it is reported).

Second, and perhaps more crucially, is the difference between belief and other states such as desire and perception. The data suggest that even these young children can report some mental states that are just immediately past. The poor performance for beliefs therefore cannot be simply a problem of poor memory or lost access. This finding presents a paradox for simulation accounts. If reporting these immediately past states requires simulation, then 3-year-olds are perfectly good simulators of their past desires and perceptions: why not beliefs? If reporting past states does not require simulation, because these states are just read off, then why do the 3-year-olds have so much trouble reporting past beliefs?

In essence, children find some sorts of mental state attributions to be difficult and some to be easy. But the difference between the easy and hard attributions is not clearly related to the distinction between self and other, as expected from ST. The distinction is related to the ability to conceive of and interpret some types of mental states and not others, for self and for other. From a theory point of view this makes sense. Even your own mental states come in several conceptual varieties, such as beliefs, desires, and perceptions, and you could be correct at reporting one variety and erroneous at another depending on your conceptual understanding of that state.

A second difficulty concerns whether children are at first generally egocentric about the mind and then overcome this by learning they must adjust their simulations for others. In Gordon's terms, is there evidence for a stage of early 'total projection'? The developmental data do not fit this general mold; there is evidence for non-egocentric understanding quite early for some states. We have already described one such task, the early 'level-1' perspective-taking task, in which children can predict that the other child will not see what they see themselves. Similarly quite young children can predict that someone else will have a desire different from their own (Wellman & Woolley, 1990). One issue for simulation theory therefore must be to explain why children who can obviously 'adjust their simulations' for some states do not do so for others, say

beliefs. Indeed, even for belief itself, the data do not suggest that children's main difficulty involves misattributing their own beliefs to others. Instead, it involves a failure to understand that beliefs can misrepresent.

This is only one example of many results that suggest that young children's errors at understanding the mind are not properly termed 'egocentric'. Even very young children are quite able to attribute to others mental states different from their own. Instead, they err by sometimes misunderstanding what certain mental states are really like.

A third empirical problem is that the simulation theory has difficulty explaining the structure of the explanations that children offer. It is commonplace to say that the child's theory is not, of course, an explicit theory but rather an implicit one, which may have to be inferred from behavior rather than being openly stated. However, in examining children's natural language and particularly their explanations for aberrant actions, we can see many explicit explanatory appeals to beliefs and desires and relations between them. One example comes from open-ended explanation tasks. In these (Wellman & Bartsch, 1989; Wellman & Banerjee, 1991) children are simply presented an action or reaction ('Jane is looking for her kitty under the piano') and asked to explain it ('Why is she doing that?'). Consider a task in which the child is asked to explain why Jane is looking for the kitty. In such an actual situation the child herself would be and should be experiencing many mental states—a fear that the kitten is lost, a creak in her back from bending down, a sensation that it is dark and not very visible under the piano, a fear the kitty will scratch, a belief the kitty is under there, a desire to find the kitty, a fantasy the kitty is a small tiger, and more. Yet children's answers to such open-ended questions are organized predominantly around beliefs and desires just as adults' are. On a simulation account why would the child answer with beliefs and desires more than fears and fantasies, pains and sensations or any of a vast number of experientially available mental states? On a simulation account there is no principled reason for the child to organize mental experiences into beliefs and desires and report those appropriately. Other empirical categories seem more compelling for categorizing and reporting first-hand mental experience (*e.g.* pains and sensations). On a theory-theory account, in contrast, there is a good reason why such explanations predominantly appeal to beliefs and desires. These are the theoretical constructs that structure the child's understanding of mental states.

More important is the shift in explanatory type between two and five, to which we have already referred. Two-year-olds' explanations almost always mention desires, but never beliefs. Asked why the girl looks for her doll under the bed they will talk about the fact that she wants the doll, but not the fact that she believes the doll is there. Three-year-olds invoke beliefs and desires, and some threes and most 4- and 5-year-olds consistently refer to the representational character of these states, explaining failure in terms of falsity. These same trends can be seen in the explanations children give in their spontaneous speech (Bartsch & Wellman, 1990).

From a ST point of view, the child's own mind, even at the very youngest ages, is a device that itself contains states like beliefs as well as desires. The child's model outputs both beliefs and desires. Why should children's explanations and predictions first privilege desires over beliefs? There is no reason to expect this if the child is simply running simulations and reporting their outcomes. From TT there is a good reason why children's explanations and predictions at first ignore beliefs and especially false beliefs or misrepresentations. Young children have yet to come to a theoretical conception of belief as an explanatory psychological construct.

A fourth difficulty involves the predictive scope of the simulation theory *versus* the theory theory. The simulation theory provides a good account of one particular type of deficit, perspective-taking difficulties, when they occur (although as mentioned earlier ST seems to mischaracterize the nature and the developmental progression of egocentric errors). However, ST fails to account for other related difficulties. For example, we (Gopnik & Graf, 1988) investigated children's ability to identify the sources of their beliefs, elaborating on a question first posed by Wimmer, Hogrefe and Perner (1988). As noted in Goldman's and Stich and Nichols' papers, there was originally some evidence suggesting that children had difficulty understanding how perceptual access leads to knowledge. More recently, however, other studies have suggested that children can indeed understand that people who see an object will know about it, while those who do not see the object will not. However, there still appear to be important limits on children's understanding of sources. For example, O'Neill, Flavell and Astington (in press) found that three-year-old children could not differentiate which source a particular piece of information might come from. They claimed for example that someone who had simply felt an object would know its color, or someone who had seen an object would know its weight.

In our experiments, we tested children's understanding of the sources of their own beliefs. Children found out about objects that were placed in a drawer in one of three ways, either they saw the objects, they were told about them, or they figured them out from a simple clue. Then we asked 'What's in the drawer?' and all the children answered correctly. Immediately after this question we asked about the source of the child's knowledge 'How do you know there's an x in the drawer? Did you see it, did I tell you about it, or did you figure it out from a clue?'. Again three-year-olds made frequent errors on this task. While they knew what the objects were, they could not say how they knew. They might say, for example, that we had told them about an object when they had actually seen it. Their performance was at better than chance levels, but was still significantly worse than the performance of four-year-olds, who were near ceiling. In a follow-up experiment (O'Neill & Gopnik, 1991) we added a condition with different and simpler source contrasts (tell, see and feel) and presented children with only two alternative possibilities at a time. We also included a control task which ensured that the children understood the meaning of

'tell', 'see' and 'feel'. Despite these simplifications of the task, the performance of the three-year-olds was similar to their performance in the original experiment. These experiments provide another striking example of the child's failure to accurately report his own mental states when they conflict with his theoretical preconceptions, and of the parallels between attributions to the self and to others.

Similarly, there is evidence for deficits in children's understanding of subjective probability. Moore *et al.* (1990) found that three-year-olds were unable to determine that a person who knew about an object was a more reliable source of information than one who merely guessed or thought. Similarly, three-year-olds, in contrast to four-year-olds, showed no preference for getting information from someone who was certain they knew what was in a box rather than someone who expressed uncertainty about their knowledge. These children seemed to divide cognitive states into full knowledge or total ignorance, they did not appreciate that belief could admit of degrees.

We believe that understanding sources and subjective probability is difficult for young children because these notions involve an understanding of the causal structure of the representational system. These aspects of the mind are not particularly different for the child and the other. However, they do require a complex causal account of the origins of beliefs. This account is at the heart of the causal-explanatory framework that eventually allows children to fully understand the representational character of the mind. These tasks should be difficult if children have not yet worked out a representational theory of mind, as we suggest, and thus should be related in development to false-belief errors. ST offers no explanation for their appearance or their relation to false-belief errors. Understanding sources and subjective probability does not seem to require complex simulation abilities, especially not when the child's own states are being reported.

In sum, the developmental pattern of children's errors and accuracies is not consistent with the view that the outputs of your own mind are simply and directly accessible, and that these outputs are attributed to others through a process of simulation. If such an account were correct, children's errors should differ between self and other in some clear fashion over development. Instead the errors divide between certain theoretical construals of inner mental states, such as beliefs *versus* desires, for both the self and the other. The child's understanding of mind is filtered through a coherent conceptual understanding of the mind; a theory. The theory organizes their interpretation of the phenomena of mental life and provides a causal-explanatory understanding of how the world informs the mind and mind guides behavior.

5. Precocity and Theory Formation

We would like to end by considering an argument that Gordon, Goldman, and Stich and Nichols share. This is the claim that children's folk psychological abilities are intellectually precocious. Children could not develop an elaborate psychological theory in a mere three or four years. Gordon and Goldman use this as an argument for simulation; children need not develop a theory of the mind, they only need to develop a mind, and run simulations on it. Stich and Nichols reply that this is an indication that important aspects of the theory are innate. We think the assumptions behind both of these arguments are ill-founded. In particular they rest on the idea that we have some *a priori* way of measuring the temporal course of conceptual change, of saying what is slow or fast or easy or difficult.

Even in the case of scientific theory change, this seems a dubious claim. How long does it take to make a theory? If we measure change sociologically it may, of course, take years or even centuries. But how long does an individual theory change take? How long did Kepler take to formulate the heliocentric theory? How long does it take a current-day student immersed in a culture that has assimilated heliocentrism to appreciate and internalize it? Days, weeks, months?

Claiming that three or four years is insufficient time for substantial theory development seems even more dubious when we consider the general cognitive achievements of young children. Developing a theory of mind is indeed an impressive achievement, but it may seem less unique if one considers parallel developments in a variety of domains. While there may be innate abilities that play a role in these achievements, there is much evidence that a great deal of abstract and complex knowledge is also learned in this period. For example, no matter how powerful the universal constraints on grammar may be, there is still an enormous amount of language-specific structure that varies sharply from one language to another. Young children quickly master these language-specific principles as well as manifesting mastery of universals (Slobin, 1981; Maratsos, 1983). More relevantly to the present case, children acquire large amounts of physical knowledge in this period. While some aspects of children's 'folk physics' are innately given, others, such as their appreciation of gravity and support, appear to be learned in months or weeks, even during infancy itself (Spelke, 1991). By four or five children also seem to have an initial understanding of biological kinds. They recognize, for example, that membership in such a kind depends on an animal's internal state, and even on its reproductive potential (Gelman & Coley, 1992).

These achievements are certainly impressive. But as we consider them it is well to remember the general intensity of the child's cognitive life. Naturalistic language data, for example, suggest that the three-year-old child may be working on the theory of mind virtually all his waking hours. And quite possibly many of his sleeping ones as well. Who knows what adults could accomplish in three years of similarly concentrated intellectual labor?

It is certainly true that there are some innately given kinds of psychological knowledge. However, it seems to us that these are most likely to be 'starting state' theories, initial conceptions of the mind that are themselves subject to radical revision in the face of evidence. They do not function as constraints on the final possibilities, in the way that, say, a Chomskyan account would propose. Moreover, it seems very unlikely that we can determine, *a priori*, which aspects of psychological knowledge are likely to be innate and which are likely to be learned. Children, for example, seem to start out as mentalists, though they must learn to be representationalists.

The evidence of developmental psychology, and indeed the evidence of common observation, suggests that young children have learning capacities (and we would claim theory formation abilities) far in excess of anything we might imagine in our daily cognitively stodgy experience as adults. Indeed we would say, not that children are little scientists but that scientists are big, and relatively slow, children. The historical progress of science is based on cognitive abilities that are first seen in very young children.

We might end by telling an evolutionary just-so story to this effect. The long immaturity of human children is a notable and distinctive feature of human beings. It seems plausible that the cognitive plasticity that is also characteristic of human beings is related to this immaturity. Human beings, unlike other species, have unique cognitive capacities to adjust their behavior to what they find out about the world. A long period of protected immaturity, the story might go, plus powerful theory-formation abilities, enable children to learn about the specific cultural and physical features of their world. These capacities typically go into abeyance once ordinary adults have learned most of what they need to know. Still, their continued existence makes specialized scientific investigation possible. Science, on this view, might be a sort of spandrel, parasitic on cognitive development itself. Young children may not only really be theorizers, they may well be better ones than we are.

Department of Psychology
University of California, Berkeley
Berkeley, CA 94720

Centre for Human Growth and Development
University of Michigan
Ann Arbor
MI 48109

References

Astington, J. W. and Gopnik, A. 1991a: Developing Understanding of Desire and Intention. In A. Whiten (eds), *Natural Theories of Mind*. Oxford: Basil Blackwell, 39–50.

Astington, J. W. and Gopnik, A. 1991b: Theoretical Explanations of Children's Understanding of the Mind. *British Journal of Developmental Psychology*, 9, 7–31.

Avis, J. and Harris, P. L. 1991: Belief-desire Reasoning Among Baka Children. *Child Development*, 62, 460–67.

Bartsch, K. and Wellman, H.M. 1989: Young Children's Attribution of Action to Beliefs and Desires. *Child Development*, 60, 946–64.

Bartsch, K. and Wellman, H. M. 1990: Everyday Talk About Beliefs and Desires: Evidence of Children's Developing Theory of Mind. Paper presented at the meeting of the Piaget Society, Philadelphia, PA.

Bretherton, I. and Beeghly, M. 1982: Talking About Internal States: The Acquisition of an Explicit Theory of Mind. *Developmental Psychology*, 18, 906–921.

Carey, S. 1985: *Conceptual Change in Childhood*. Cambridge, MA.: MIT Press.

Carey, S. 1988: Conceptual Differences Between Children and Adults. *Mind and Language*, 3, 167–181.

Churchland, P. M. 1984: *Matter and Consciousness*. Cambridge, MA.: MIT Press.

Dretske, F. 1981: *Knowledge and the Flow of Information*. Cambridge, MA.: MIT Press.

Estes, D., Wellman, H. M. and Woolley, J. D. 1989: Children's Understanding of Mental Phenomena. In H. Reese (ed.), *Advances in Child Development and Behavior*. New York: Academic Press, 41–87.

Flavell, J. H., Everett, B. A., Croft, K. and Flavell, E. R. 1981: Young Children's Knowledge About Visual Perception: Further Evidence for the Level 1–Level 2 Distinction. *Developmental Psychology*, 17, 99–103.

Flavell, J. H., Green, F. L. and Flavell, E. R. 1986: Development of Knowledge About the Appearance-Reality Distinction. *Monographs of the Society for Research in Child Development*, 51 (Serial No. 212).

Flavell, J. H., Flavell, E. r. and Green, F. L. 1987: Young Children's Knowledge About Apparent-Real and Pretend-real Distinctions. *Developmental Psychology*, 23, 816–22.

Flavell, J. H. 1988: The Development of Children's Knowledge About the Mind: From Cognitive Connections to Mental Representations. In J. Astington, P. Harris and D. Olson (eds.), *Developing Theories of Mind*. New York: Cambridge University Press, 244–67.

Flavell, J. H., Flavell, E. R., Green, F. L. and Moses, L. J. 1990: Young Children's Understanding of Fact Beliefs versus Value Beliefs. *Child Development*, 61, 915–28.

Forguson, L. and Gopnik, A. 1988: The Ontogeny of Common Sense. In J. Astington, P. Harris and D. Olson (eds.), *Developing Theories of Mind*. New York: Cambridge University Press, 226–43.

Gelman, S. A. and Coley, J. D. 1992: Language and Categorization: The Acquisition of Natural Kind Terms. In S. A. Gelman and J. P. Byrnes (eds.), *Perspectives on Languages and Thought*. Cambridge: Cambridge University Press.

Gopnik, A. 1984: Conceptual and Semantic Change in Scientists and Children: Why There Are No Semantic Universals. *Linguistics*, 20, 163–79.

Gopnik, A. 1988: Conceptual and Semantic Development as Theory Change. *Mind and Language*, 3, 197–217.

Gopnik, A. and Astington, J. W. 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. and Graf, P. 1988: Knowing How You Know: Young Children's Ability to Identify and Remember the Sources of Their Beliefs. *Child Development*, 59, 1366–71.

Gopnik, A. and Slaughter, V. 1991: Young Children's Understanding of Changes in Their Mental States. *Child Development*, 62, 98–110.

Gopnik, A. and Slaughter, V. 1992: Children's Understanding of Perception and Belief. Unpublished manuscript.

Harris, P. L. 1991: The Work of the Imagination. In A. Whiten (ed.), *Natural Theories of Mind*. Oxford: Basil Blackwell, 283–304.

Harris, P. L., Brown, E., Marriot, C., Whittal, S. and Harmer, S. 1991: Monsters, Ghosts and Witches: Testing the Limits of the Fantasy-reality Distinction in Young Children. *British Journal of Developmental Psychology*, 9, 105–123.

Johnson, C. N. 1988: Theory of Mind and the Structure of Conscious Experience. In J. Astington, P. Harris and D. Olson (eds.), *Developing Theories of Mind*. New York: Cambridge University Press, 47–63.

Karmiloff-Smith, A. and Inhelder, B. 1975: If You Want to Get Ahead, Get a Theory. *Cognition*, 3, 195–212.

Keil, F. C. 1989: *Concepts, Kinds, and Cognitive Development*. Cambridge, MA.: MIT Press.

Leslie, A. M. 1987: Pretense and Representation: The Origins of 'Theory of Mind'. *Psychological Review*, 94, 412–26.

Masangkay, Z. S., McCluskey, K. A., McIntyre, C. W., Sims-Knight, J., Voughn, B. E. and Flavell, J. H. 1974: The Early Development of Inferences About the Visual Percepts of Others. *Child Development*, 45, 357–66.

Meltzoff, A. N. and Gopnik, A. In press: The Role of Imitation in Understanding Persons and Developing Theories of Mind. In S. Baron-Cohen and H. Tager-Flusberg (eds.), *The Theory of Mind Deficit in Autism*. New York: Cambridge University Press.

Meltzoff, A. N. and Moore, M. K. 1977: Imitation of Facial and Manual Gestures by Human Neonates. *Science*, 198, 75–8.

Mitchell, P. and Lacohee, H. In press: Children's Early Understanding of False Belief. *Cognition*.

Moore, C., Pure, K. and Furrow, P. 1990: Children's Understanding of the Modal Expression of Certainty and Uncertainty and its Relation to the Development of a Representational Theory of Mind. *Child Development*, 61, 722–30.

Moses, L. J. 1990: Young Children's Understanding of Intention and Belief. Unpublished Ph.D. dissertation, Stanford University.

Moses, L. J. and Flavell, J. H. 1990: Inferring False Beliefs from Actions and Reactions. *Child Development*, 61, 929–45.

O'Neill, D. K., Astington, J. W. and Flavell, J. H. In press: Young Children's Understanding of the Role that Sensory Experiences Play in Knowledge Acquisition. *Child Development*.

O'Neill, D. K. and Gopnik, A. 1991: Young Children's Ability to Identify the Sources of Their Beliefs. *Developmental Psychology*, 27, 390–99.

Perner, J., Leekam, S. R. and Wimmer, H. 1987: Three-year-olds' Difficulty with False Belief. *British Journal of Developmental Psychology*, 5, 125–37.

Perner, J. 1991: *Understanding the Representational Mind.* Cambridge, MA.: MIT Press.

Searle, J. R. 1983: *Intentionality.* New York: Cambridge University Press.

Shatz, M., Wellman, H. M. and Silber, S. 1983: The Acquisition of Mental Verbs: A Systematic Investigation of First References to Mental State. *Cognition,* 14, 301–321.

Siegal, M. and Beattie, K. 1991: Where to Look First for Children's Understanding of False Beliefs. *Cognition,* 38, 1–12.

Slobin, D. I. 1981: The Origin of Grammatical Encoding of Events. In W. Deutsch (ed.), *The Child's Construction of Language.* New York: Academic Press.

Spelke, E. S. 1991: Physical Knowledge in Infancy. In S. C. and R. Gelman (eds.), *The Epigenesis of Mind: Essays on Biology and Cognition.* Hillsdale NJ.: Lawrence Erlbaum Associates, 133–69.

Stich, S. 1983: *From Folk Psychology to Cognitive Science.* Cambridge, MA.: MIT Press.

Trevarthen, C. and Hubley, P. 1978: Secondary Intersubjectivity: Confidence, Confiders, and Acts of Meaning in the First Year of Life. In A. Lock (ed.), *Before Speech: The Beginning of Interpersonal Communication.* New York: Academic Press.

Wellman, H. W. In press: Early Understanding of Mind: The Normal Case. In S. Baron-Cohen, H. Tager-Flusberg, Cohen and Volkman (eds.), *Understanding Other Minds: Perspectives From Autism.* Oxford: Oxford University Press.

Wellman, H. M. and Estes, D. 1986: Early Understanding of Mental Entities: A Reexamination of Childhood Realism. *Child Development,* 57, 910–23.

Wellman, H. M. and Bartsch, K. 1988: Young Children's Reasoning About Beliefs. *Cognition,* 30, 239–77.

Wellman, H. M. and Gelman, S. 1988: Children's Understanding of the Nonobvious. In R. J. Sternberg (ed.), *Advances in the psychology of intelligence Volume 4.* Hillsdale, NJ.: Lawrence Erlbaum Associates.

Wellman, H. M. 1990: *The Child's Theory of Mind.* Cambridge MA.: MIT Press.

Wellman, H. M. and Woolley, J. D. 1990: From Simple Desires to Ordinary Beliefs: The Early Development of Everyday Psychology. *Cognition,* 35, 245–75.

Wellman, H. M. and Banerjee, M. 1991: Mind and Emotion: Children's Understanding of the Emotional Consequences of Beliefs and Desires. *British Journal of Developmental Psychology,* 9, 191–224.

Wellman, H.M. and Gelman, S. A. In press: Cognitive Development: Foundational Theories of Core Domains. *Annual Review of Psychology.*

Wimmer, H. and Perner, J. 1983: Beliefs About Beliefs: Representation and Constraining Function of Wrong Beliefs in Young Children's Understanding of Deception. *Cognition,* 13, 103–128.

Wimmer, H., Hogrefe, J. and Perner, J. 1988: Children's Understanding of Informational Access as Source of Knowledge. *Child Development,* 59, 386–96.

Wimmer, H. and Hartl, M. 1991: Against the Cartesian View on Mind: Young Children's Difficulty with Own False Beliefs. *British Journal of Developmental Psychology,* 9, 125–38.

Yuill, N. 1984: Young Children's Coordination of Motive and Outcome in Judgments of Satisfaction and Morality. *British Journal of Developmental Psychology,* 2, 73–81.

Zaitchek, D. 1990: When Representations Conflict With Reality: The Preschooler's Problem with False Beliefs and 'False' Photographs. *Cognition,* 35, 41–68.