Theory-of-Mind Development: Retrospect and Prospect

John H. Flavell, Stanford University

This review begins with a brief history from Piagetian perspective-taking development, through metacognitive development, and into the past and present field of theory-of-mind development. This field has included research on what infants and children know about a variety of mental states, on possible causes and consequences of mentalistic knowledge, and on similarities and differences in this knowledge across individuals, cultures, and primate species. The article concludes with some speculations about the future of the field.

Theory-of-mind development is the area of cognitive development research that investigates the nature and development of our understanding of the mental world—the inner world inhabited by beliefs, desires, emotions, thoughts, perceptions, intentions, and other mental states. Since its beginnings some 20 years ago, this area has grown to be one of the largest and liveliest in development psychology. As one indication of this area’s size, as long ago as 1998 Wellman, Cross, and Watson (2001) were able to identify some 178 studies in just the subarea of children’s false-belief understanding. As another indication, a recent computer search turned up 399 publications containing the phrase theory of mind. Research on theory of mind has proven to be of interest, not only to developmental psychologists but also to researchers and practitioners in fields such as philosophy, psychiatry, neuropsychology, social psychology, clinical psychology, comparative psychology, cultural psychology, cognitive psychology, and education. Reviews of theory-of-mind development research include Baron-Cohen, Tager-Flusberg, and Cohen (2000), Bartsch and Wellman (1995), Flavell...
The article begins with a brief history of the area, including a summary of past and present research directions. It concludes with some predictions about the future of the area. In the interest of brevity and flow and because this area has been thoroughly reviewed previously, I will be fairly sparing of references. Thus, the article is intended to be more a high-altitude overview of the area than a thoroughly detailed review of it.

Retrospect

As is true of so many areas of cognitive development, the history of this one mainly begins with Piaget (Flavell, 2000; Flavell & Miller, 1998; Shantz, 1983). A central Piagetian claim was that children begin development by being cognitively egocentric. Piaget and his colleagues used egocentrism and other concepts to interpret their developmental studies of a wide variety of social-cognitive topics: perceptual perspective-taking; egocentric communication; the misattribution of mental characteristics to physical objects (animism) and physical characteristics to mental events (realism); and understanding of thoughts, dreams, intentions, and morality. Research on some of these topics still continues, although usually not from a Piagetian theoretical perspective (e.g., Flavell, Green, & Flavell, 1995b; Woolley & Boerger, 2002).

A second wave of theory and research in this general area was the extensive work on metacognitive development that began in the early 1970s. Surveys of this large literature include Brown, Bransford, Ferrara, and Campione (1983), Flavell, Miller, and Miller (2002), Kuhn (1999), Moshman (1998), and Schneider and Bjorklund (1998). Metacognition (cognition about cognition—hence the _meta_ ) has been defined as any knowledge or cognitive activity that takes as its object, or regulates, any aspect of any cognitive activity (Flavell, Miller, & Miller, 2002). The majority of developmental studies classified as metacognitive have investigated children’s metamemory—that is, their knowledge about variables affecting memory performance and, especially, their knowledge and use of memory strategies. However, the term has also been applied to numerous studies of children’s cognition concerning comprehension, communication, language, perception, attention, and problem solving. Research in the metacognitive development tradition is still being done, although it is not the hot topic it used to be. On the other hand, metacognition in adults appears to be a
thrusting field (Jost, Kruglanski, & Nelson, 1998; Yzerbyt, Lories, &
knowledge about the mind would probably classify their work as either
metacognitive or in the general Piagetian tradition. Today, most would
say they are doing one or another kind of theory-of-mind research.
What happened to bring about this change?

In the 1978 issue of *Behavioral and Brain Sciences*, Premack and
Woodruff reported some research to test whether chimpanzees had
what they called a “theory of mind.” In their commentaries on this
article, three philosophers independently suggested that one might be
able to find out whether an animal possessed the concept of belief in
something like the following fashion (Bennett, 1978; Dennett, 1978;
Harman, 1978): The subject animal sees another individual put an
object in container A and then leave the scene. The subject then sees
someone else transfer the object from container A into container B
while the individual is still absent. The subject animal should then be
credited with some understanding of belief if it acts as if it expects
that the returning individual will search for the object in A rather
than B.

These ideas were given research expression in the early 1980s by
two Austrian psychologists, Josef Perner and Heinz Wimmer. In a pio-
nearning and highly influential series of studies, they used the “unex-
pected transfer” method proposed by the philosophers to test young
children’s understanding of false belief (Wimmer & Perner, 1983).
Other false-belief tasks of the following type were also developed:
After children discover that a cookie box actually contains pencils
instead of cookies, they are asked what another child who has not
looked inside will think the box contains. Younger preschoolers say
pencils; older ones, with a better understanding of belief, say cookies.
Around the same time, Wellman and his coworkers had independently
begun to conceptualize children’s developing metacognitive knowledge
and understanding of mental terms as the development of a theory of
mind (e.g., Wellman, 1985). In addition, other researchers who had not
yet begun to conceptualize children’s development in quite this way
had been doing research that subsequently became part of the theory-
of-mind movement. An example would be the work on children’s
knowledge about perception and about the appearance-reality distinc-
tion by Flavell and colleagues (e.g., Flavell, Flavell, Green, & Wilcox,
1980; Flavell, Flavell, & Green, 1983; see Astington, Harris, & Olsen,
1988, for other such projects).

The movement was given added identity and coherence by two
conferences that were held in the spring of 1986. The presentations
given at these two conferences were later published in a book entitled *Developing Theories of Mind* (Astington, Harris, & Olson, 1988), and the movement was officially launched. Figure 1 illustrates the main directions that theory-of-mind development research has taken since it began in the early 1980s. Much of the earliest work was focused on documenting a striking improvement between 3 and 5 years of age in children’s performance on various false-belief (FB), appearance-reality (AR), and Level 2 visual perspective-taking (PT) tasks. Thus, for example, older but not younger preschoolers were usually found to show an understanding that a naive other child would falsely believe that the cookie box contains cookies (false belief), that a fake rock looks like a rock but is really a sponge (appearance-reality), and that a picture book that is oriented correctly for them on the table will look upside down to a person seated opposite (Level 2 visual PT).

From those beginnings work has progressed more or less concurrently in a variety of directions, as shown by the arrows in Figure 1 and as briefly elaborated in the sections that follow. Fuller discussion of these topics can be found in the substantive reviews cited earlier.
Theories and Antecedents

Several types of theories have been offered as explanations for the development of children's mentalistic understanding. One is the so-called theory theory (Gopnik & Meltzoff, 1997; Gopnik & Wellman, 1994; Perner, 1991; Wellman & Gelman, 1998). Theory theorists argue that our knowledge about the mind comprises not a formal scientific theory but an informal, everyday framework or foundational theory. A number of steps in children's progression toward the adult theory of mind have been described. For instance, Bartsch and Wellman (1995) have argued that children begin with a desire psychology, then progress to a desire-belief psychology, and finally attain our adult belief-desire psychology, in which one recognizes that what people believe, as well as what they desire, crucially affects how they behave. Theory theorists argue that experience plays a major formative role in children's theory-of-mind development.

In contrast, modularity theorists (Baron-Cohen, 1995; Leslie, 1994; Scholl & Leslie, 1999) postulated the acquisition through neurological maturation of a succession of domain-specific and modular mechanisms for dealing with agents versus nonagent objects. Although experience may be necessary to trigger the operation of these mechanisms, it does not determine their nature. Although obviously focused on basic hard-wired competence in this area, these therapists do not neglect performance factors influencing its expression.

Harris (1992) and others have proposed yet a third approach. According to their simulation theory, children become able to compute the mental states of other people through a kind of role-taking or simulation process. What develops is the ability to make increasingly accurate simulations of this kind. Like theory theorists, simulation theorists also assume that experience plays a crucial formative role, in that it is through practice in role taking that children improve their simulation abilities.

A number of developmentalists believe that young children's failures on false-belief and other theory-of-mind tasks may be caused by limitations in executive functioning (e.g., Carlson, Moses, & Hix, 1998; Hughes, 1998). For example, an inability to inhibit a dominant, ready-to-go response could cause the child subject to blurt out the cognitively salient real contents of the cookie box when asked what the naive other child thought it contained. According to this view, improvement in children's executive functioning with age helps make possible the acquisition of key theory-of-mind competencies.

Another important antecedent is language development. It is particularly easy to imagine how language competencies might play a variety of roles in assisting theory-of-mind development. People convey
information about their own, the child’s, or other people’s mental states through conversations and stories, verbally make salient different people’s perspectives, and help the child see how mental states are caused and changed by verbal and other inputs. Language also provides vehicles for thinking about mental states: a vocabulary of mental terms, a complementation sentence structure for expressing propositional attitudes (e.g., “She thinks that he is nice.”), and a way of abstracting and reflecting on mental states and behaviors. And, in fact, the research evidence provides good support for these conjectures about the formative influence of language and interpersonal communication (Astington & Baird, in press).

**Consequents**

A number of studies have shown that children’s theory-of-mind development has consequences for their social behavior; for recent reviews, see Repacholi and Slaughter (2003). Most research evidence suggests that children with more advanced understanding (especially of false belief) tend to have more successful social relationships than their less advanced peers. For example, they are rated by their teachers as having better social skills (Watson, Nixon, Wilson, & Capage, 1999). However, there is also evidence that theory-of-mind competencies can be used for antisocial as well as presocial purposes. For example, some child bullies and adult sociopaths consistently employ their mind-reading skills to serve antisocial ends (Repacholi & Slaughter, 2003). Finally, the causal relations between theory of mind and social behavior are complex and bidirectional, with social behavior providing a context for theory-of-mind acquisitions as well as the converse (Astington, 2003). Thus, the two could be viewed as both antecedents and consequents of one another.

**Developments During Infancy**

Babies are born with or early acquire a number of abilities and propensities that will help them learn about people. They find human faces, voices, and movements especially interesting. They seem impelled to attend to and interact with other people, and they certainly impel other people to attend to and interact with them. Infants respond differently to people than they do to objects and seem to expect people to behave differently than objects do (Poulin-Dubois, 1999). They appear to construe people as agents that are self-propelled, goal directed, and influenceable at a distance by communicative signals. All of these seem like the right design features for a creature destined for theory-of-mind development.
Late in the first year, perhaps earlier, infants are beginning to learn that people’s behavior possesses intentionality, or “aboutness” (see Tomasello & Haberl, 2003, for a good, up-to-date review of this general topic). An individual's behavior is *about* an object in this sense if the individual perceptually attends to it, labels it, thinks about it, wants it, fears it, intends or tries to get it, or relates to it in any other psychological way. Infants do a variety of things that reflect a beginning awareness of intentionality. They try to engender new aboutnesses in others through various communicative gestures, such as pointing to or vocalizing about an object and checking to see if the other person attends to it. They also develop skill at reading the aboutnesses the other person is already displaying, as when they follow the person’s gaze. Carpenter, Nagell, and Tomasello (1998) have recently documented a three-step developmental sequence in which infants progress from *sharing* to *following* to *directing* others’ attention and behavior. Studies by Meltzoff (1995) have demonstrated that 18-month-olds can infer what action another person is trying to perform (e.g., attempting to pull one object away from another object to which it is attached), even though the person is unsuccessful in the attempt (does not succeed in pulling it away) and therefore never actually demonstrates the intended outcome. This and other findings suggest that infants of this age have some sense that people’s actions are intentional and goal-directed. By age 18 months infants also understand that they should give an experimenter a food that she reacts to with pleasure rather than one toward which she acts disgusted, even when they themselves prefer the latter food (Repacholi & Gopnik, 1997); this suggests at least some limited ability to reason nonegocentrically about people’s desires.

Infants also recognize that it is the adult’s intentional focus rather than their own that gives clues as to the adult’s referential intent when the adult labels an object (Woodward & Markman, 1998). That is, they recognize that the word refers to the object the adult is currently looking at, not the one they themselves happened to be looking at when the word was spoken. In this way, early theory-of-mind development provides a scaffold for early language development. Similarly, babies develop the ability to learn what an object is like by reading the adult’s attentional focus when the adult is expressing a positive or negative emotional reaction to it (a process called *social referencing*). For instance, they may selectively avoid an object toward which their parent shows negative affect. Thus, by 12 months or so they can recognize that the adult’s emotional display refers to, or is about, a particular object much as they can recognize that the adult’s spoken label refers
to, or is about, a particular object (Moses, Baldwin, Rosicky, & Tidball, 2001). By the end of infancy children may also do other things suggestive of a beginning understanding of human psychology, such as trying to comfort people in distress and correctly using mental state terms such as want and see (Flavell & Miller, 1998).

**Later Developments**

A very large literature has accrued since the early 1980s on theory-of-mind acquisitions that occur subsequent to the infancy period. What follows is a very brief summary of some of the major findings, organized by type of mental state.

**Visual Perception.** During the early preschool period children already realize that a person will see an object if and only if the person's eyes are aimed in the general direction of the object, and if there are no vision-blocking obstacles interposed between the person and the object (Flavell, 1992, in press). With this understanding, they are able to do simple, nonegocentric visual perspective-taking; for example, they can infer that you may see something that they do not and vice versa (referred to as Level 1 knowledge about visual perception). Later in the preschool period they go on to recognize that the same thing may present different visual appearances to two people if they view it from different positions (called Level 2 knowledge).

**Attention.** As already mentioned, even infants pay attention to other people's attending and seem to have some understanding of its implications (Tomasello & Haberl, 2003). In subsequent years they come to appreciate that attention is selective and limited and that different people may mentally represent the same attended-to input differently (Fabricius & Schwanenflugel, 1994; Flavell, Green, & Flavell, 1995a; Pillow, 1995).

**Desires.** By the age of 3 children are not only using some desire terms correctly, they also seem to grasp simple causal relations among desires, outcomes, emotions, and actions—suggestive evidence that they are developing something like an implicit theory. For example, they understand that people will feel good if they get what they want and feel bad if they do not (Bartsch & Wellman, 1995).

**Emotions.** Although we do not know yet whether infants actually attribute inner feelings to people who display emotions, it seems certain that young preschoolers do (Wellman, Harris, Banerjee, & Sinclair, 1995). In later years children learn more advanced truths about emotions—for example, that people do not always really feel what they appear to feel and that people's emotional reactions to an event may be influenced by earlier emotional experiences with similar events or by their current mood (Flavell & Miller, 1998).
Beliefs and Related Mental Representations. There have been a great many studies of children’s developing understanding of so-called serious mental representations—that is, nonpretense mental states, such as beliefs that are meant to represent reality accurately (Flavell & Miller, 1998). The majority of these studies have dealt with children’s comprehension of representations that differ from person to person or differ from reality: the appearance-reality distinction, Level 2 knowledge of visual perception, interpretation and constructive processing, deception, and most studied of all, false belief. How well standard false-belief and appearance-reality tasks actually measure these concepts is currently controversial, however (e.g., Bloom & German, 2000; Hansen, 2003). Children’s knowledge about mental representations continues to increase after the preschool period. In particular, it is not until middle childhood and later that children appear to gain any substantial understanding of the mind as an active, interpretive, constructive processor (e.g., Barquero, Robinson, & Thomas, 2003; Carpendale & Chandler, 1996). For instance, understanding that people’s interpretation of an ambiguous event may be influenced by their preexisting biases or expectations seems to be a largely middle-childhood insight (Pillow & Henrichon, 1996).

Knowledge. Young preschoolers appear to be unclear about just what it means for someone to know something and about how knowledge is acquired (Flavell & Miller, 1998). Even older preschoolers may claim that they have always known information that they have just learned during the experimental session (Taylor, Esbensen, & Bennett, 1994). An important early middle-childhood discovery is that perceptual information has to be adequate as well as merely present to engender knowledge. For example, children come to realize that a person often cannot be certain of an object’s identity when only a little bit of it is visible; this realization is another example of their burgeoning conception of the mind as an interpretive device.

Pretense. Leslie (1987, 1994) has argued that the ability to understand pretense and the ability to understand false belief and other mental states are mediated by a common, early-maturing theory-of-mind module. This argument has some plausibility: “Pretending that” and “believing that” are both propositional attitudes. Moreover, adults regard both as mental representations or construals of something as being a certain way—either for real (belief) or just temporarily, for play purposes (pretense). Nevertheless, Leslie’s claim is currently controversial (Harris, 2000; Lillard, 1998a). The related topic of children’s understanding of imagination is also being studied (Harris, 2000; Woolley, 1995).
Thinking. Children achieve some important elementary knowledge and skills concerning thinking during the early preschool years (Flavell, Green, & Flavell, 1995b; Wellman, Hollander, & Schult, 1996). For example, they come to construe thinking it as an internal human activity that refers to or represents real or imaginary things. However, there are also important knowledge and skills concerning thinking that preschoolers clearly lack (Flavell, 2003; Flavell, Green, & Flavell, 1995b; Flavell & O’Donnell, 1999). They are not aware that people are continually experiencing mental content spontaneously in an ever-flowing stream of consciousness. For example, unlike older children, preschoolers do not consistently attribute any mental activity at all to a person who just sits quietly, “waiting.”

Differences in Development

Intracultural Differences. Investigators have examined three kinds of differences in development: intracultural, intercultural, and interspecies (Flavell & Miller, 1998). Regarding intracultural differences, the previous section on theories and antecedents mentioned the importance of social-communicative experiences as mediators of within-culture differences in theory-of-mind development (Astoning & Baird, in press; Repacholi & Slaughter, 2003). The most striking intracultural differences, however, are seen in the pronounced deficits in theory-of-mind development in autistic individuals.

Significant intracultural variation that does not necessarily imply deficits or incomplete development also exists. Dweck and her coworkers have documented important individual differences in people’s implicit theories about intelligence and other human attributes (Dweck, Chin, & Hong, 1995). Textbooks in the fields of personality, social psychology, and social cognition also describe many other ways that normal adults differ from one another in their naive theories and knowledge regarding themselves and other people; great works of literature are an even richer source. And of course psychologists and other scientists have espoused widely different conceptions of human cognition and personality over the years: Think, for example, of the differences between B. F. Skinner’s and Sigmund Freud’s views of the mind.

Intercultural Differences. Researchers have also begun to ask questions about the extent to which theory-of-mind development is similar across cultures. It seems likely that at least the fundamental mentalistic understanding observed in infants and very young children would be found universally (Wellman, 1998). There is even evidence that some developmental sequences may be cross-culturally invariant: In a careful study, Tardif and Wellman (2000) showed that both Chinese and American children acquire basic desire and belief concepts and
also that both acquire them in that order. Some cross-cultural variation obtains in later development, however (Greenfield et al., 2003; Lillard, 1998b; Vinden & Astington, 2000). For example, some languages encode mental states more richly than others do, and some cultures (and subcultures) encourage mentalistic thought and talk more than others do.

**Interspecies Differences.** There is a lively controversy in the recent comparative psychology literature as to whether other primates—especially chimpanzees—should be credited with any genuinely mentalistic knowledge. Povinelli and colleagues (Povinelli & Vonk, 2003) believe that they are largely lacking in such knowledge: For example, these researchers have offered evidence suggesting that chimps possess a behavioristic rather than a mentalistic conception of seeing. In contrast, Tomasello and colleagues (Tomasello, Call, & Hare, 2003a, 2003b) argue that chimps do understand some psychological states and that “the question is only which ones and to what extent” (Tomasello et al., 2003a, p. 156). Although both labs have done very clever experiments to address the question, exactly what nonhuman primates do and do not understand in the mental domain remains unclear. That the question is an important one, however, is underscored by this comment from Tomasello et al.: “At issue is no less than the nature of human cognitive uniqueness” (2003a, p. 156).

**Prospect**

Researchers have learned a great deal about theory-of-mind development in the past two decades. It has proven to be an immensely fruitful area of developmental psychological study. Will it continue to be so in the next few decades? It is hard to see why it would not. There are a number of outstanding questions that future research will continue to address (Flavell, 2000). As one example, what factors, processes, or mechanisms (e.g., language, executive functions) contribute to children’s theory-of-mind development and exactly what is the nature of their contributions? As another, how much and what kinds of intracultural, intercultural, and interspecies similarities and differences exist among adults in this area?

One exciting future prospect is the possibility of telling longer and richer developmental stories in key sectors of theory-of-mind development (Flavell, 2000). Take, for example, the sector of belief and knowledge development. There is more to the developmental story here than that 4-year-olds tend to pass standard false-belief and knowledge tasks and 3-year-olds tend to fail them. Continuing the story forward, many older children and adults will progress to other discoveries: as exam-
ple, that the mind is an interpretive and constructive device (Carpendale & Chandler, 1996), that there can be higher-order as well as first-order beliefs (e.g., “He thinks that she thinks.”), and that scientific and other advanced forms of reasoning and knowledge acquisition have special rules (Kitchener, 2002; Kuhn, 2000; Moshman, 1998). Research on individual differences in adulthood might even include the study of virtuosos in this area—writers, therapists, or others who have acquired exceptional perspective-taking or introspective skills (Gardner, 1983).

As to earlier competencies in this same sector, exciting research is currently being done on what may be some implicit understanding of belief and knowledge in infants and very young children. Several studies indicate that young 3-year-olds who respond incorrectly to standard false-belief task questions nonetheless show by their eye movements that they may have some implicit understanding of false beliefs (Clements & Perner, 1994; Garnham & Ruffman, 2001). Even more surprising, Onishi and Baillargeon (2002) have claimed evidence for such understanding in 15-month-olds. For comparable findings regarding a possible implicit grasp of the concept of knowledge, see O’Neill (1996) and Tomasello and Haberl (2003). Knowledge about vision (Flavell, in press) and about intentions (Schult, 2003; Tomasello & Haberl, 2003) also show extended developmental itineraries.

Finally, consideration of extended developmental itineraries raises deep questions about developmental diagnosis (Flavell, Miller, & Miller, 2002). How should we conceptualize the various items of mentalistic knowledge that individuals acquire in the course of ontogenesis? Is a given item best conceptualized as implicit, explicit, or some mix of the two? And if explicit, how often is the individual consciously aware of it and how able and disposed to reflect on it? What is the individual capable of using the knowledge for, and how often does the individual actually put it to that use in everyday life? In short, rather like the developing child, we need to learn the different things that it can mean for a creature to “know” something.

References


SCHULT, C. A. (2003). Wanting this but planning that: Children’s understanding of intentions and desires. Unpublished manuscript, Indiana University, South Bend.


