The Development of Theory of Mind: Historical Reflections

Henry M. Wellman
University of Michigan

ABSTRACT—In this article, I reflect on theory of mind as a field, including how it arose and how it developed. My research has been intertwined with this process: beginning right out of graduate school, my career developed along with the field, and I contributed to the field and its development at various points. So this essay also traces my path as I strived, and still strive, to understand how theory of mind begins and unfolds in human development, what forces shape that development, and what accounts best explain the timetables and progressions of theory-of-mind understandings in humans. I end with my sense of where theory-of-mind research is likely to head in the near future.

KEYWORDS—history; theory of mind

Edmund Wilson, the famous evolutionary biologist, reportedly told aspiring researchers that a good scientist should be bright enough to spot a promising research endeavor, but not so bright as to become bored doing it. For 35 years, I have not been bored with theory of mind.

There was a time before theory of mind, of course, both for developmental science and for me. For graduate school, I chose the Institute of Child Development at Minnesota because, being a preschool teacher, there I could opt for either an academic or an early childhood education degree. Fortunately, my assigned advisor was John Flavell. He was the foremost expert on Piaget outside of Geneva, but at that point I had not heard of Flavell or Piaget. I was quickly drawn to Piaget for his insistence on constructionism to understand development, an approach I believe in to this day. But my research with Flavell was focused on memory development and then metacognition, a topic that was being created from scratch in my years at Minnesota.

My interest in metacognition morphed into a more basic one involving what children think about the mind more generally, not just their ideas about memory and learning. This more aptly fed my Piaget-fueled intrigue with children's concepts. At Minnesota, Carl Johnson and I began revisiting Piaget's claims about children's understanding of mind.

One thing that benefitted the field and me is that, early on, theory of mind attracted some unusually insightful scholars who, while often at odds, were collegial and open to argument, data, and alternative viewpoints; these included Paul Harris, Alison Gopnik, Josef Perner, Alan Leslie, and Janet Astington. They and others continued developing the topic while recruiting students and junior colleagues to the endeavor, and seducing some senior scholars like Flavell.

In what follows, I outline where we have been, where we are, and my sense of directions for the future. My own insider appraisals clearly color this synopsis: Mine is not an unbiased perspective and this is not unbiased history.

ORIGINS, IN TWOs

The question of how people come to understand their own and others' minds has a long history in philosophy and psychology. But two thrusts launched the field within developmental science. For some, like me, interests morphed from metacognition to mind. For others, David Premack and Guy Woodruff's (1) seminal article, “Does the chimpanzee have a theory of mind?” set things off. That phrase—theory of mind—caught on quickly, in part, because it was catchy. But for some of us, it also aptly fit an emerging theory theory: the theoretical claim that children's conceptual development constitutes naïve theory development (2–4). In my view, theory of mind remains the best example of a foundational theory of everyday cognition, an early developing naïve psychology that complements children’s naïve physics and biology (5).
Two conferences in 1986 incubated the field further, one organized by Astington and Gopnik at Toronto, one organized by Harris at Oxford. Several of us attended both; the resulting papers were collected into a book (6), a rare edited volume that shaped research for years to come. Two early books, one by me (7) and one by Perner (8), helped create further interest.

Finally, two aspects of mind were covered in early work: roughly, mental states and mental entities. Understanding mental states was studied by assessing 3- to 6-year-olds’ appreciation of how agents’ beliefs and desires work together to produce intentional behavior (9), including actions driven by false beliefs (10). Comprehending mental entities involved assessing children’s understanding that thoughts and physical objects (e.g., a thought about a dog vs. a dog, or thoughts vs. close imposters, such as air, shadows, and moving pictures) are two very different sorts of things, one nonmaterial and mental, the other physical, tangible, and real (7). Contradicting Piaget’s claims about childhood realism, young children were surprisingly good at both of these nonobvious and abstract understandings.

**DEVELOPMENT**

Development includes origins and change, and so does this field. Over the years, research expanded in breadth and depth, producing studies on the development of theory of mind in infancy, childhood, adolescence, adulthood, and even late life; on emotion understanding; and on children’s conversations about people and minds through the use of terms like think, want, feel, and guess. Researchers have also examined underpinnings of theory-of-mind understandings in evolutionary processes, neural processes, and developmental learning (including computational models of constructivist learning); development in typical and atypical populations (e.g., individuals with autism or deafness); and individual differences in theory-of-mind development.

Individual differences are no more or less important than these other topics, but individual variation in reaching theory-of-mind milestones helped address antecedents of theory-of-mind competence, such as engagement in pretend play, having siblings, frequency of engagement in explanatory conversations, and growing up bilingual. The research has been not only longitudinal and correlational, but also experimental, including microgenetic studies designed to accelerate the ordinary experiences that arguably propel theory-of-mind acquisition. Similarly, such variation allowed investigators to show that theory-of-mind developments have wide-ranging consequences, including significant impacts on children’s friendships and popularity, engagement in lying and deception, game-playing skills, strategies for persuading or arguing with others, and transition to school.

The list does not end there: Theory-of-mind advances predict children’s cognitive skills, such as their metacognitive strategies (harkening back to metacognition), learning reading and mathematics content, and acceptance of feedback from teachers. Researchers have looked at variations in upbringing to identify both universal and culture-specific aspects of theory-of-mind conceptions and trajectories. They have also examined relations with languages being acquired. Research with monkeys, chimpanzees, and surprisingly, dogs has shed light on the nature of distinctively human theory of mind (harkening back to Premack and Woodruff). That research has inspired theories, such as the social brain hypothesis (11), which argue that evolutionary increases in social cognition fueled advances in human intelligence more broadly. (For supporting information and references for these areas, see ref. 12.) These discoveries attest to the health of the field. Moreover, several of these topics are particularly rich in data and debate, allowing for deeper consideration of past developments and pointing to directions for research.

**Developmental Progressions in Understanding Theory of Mind**

Early on, many researchers studying theory of mind became obsessed with false belief. False-belief tasks, as in Figure 1, can be very simple, and have many variants appropriate to different cultures and contexts. This led to hundreds of studies and several revealing meta-analyses e.g., (13). While false belief was, in hindsight, a lucky find, focus on a single task (even in aggregated batteries) is misleading and “not very developmental” (14, p. 316). A wider developmental perspective was clearly needed.

Theory-of-mind understandings begin in infancy, but also progress: Early understandings of intentional action give way to later belief-desire systems of understanding, where children show knowledge that actions are produced from an agent’s desires and beliefs in combination. At its simplest, we see people engaging in actions they believe will get them what they want. Even preschoolers do this.

Clearly, such understandings must encompass notions about desires as well as beliefs, and researchers discovered early that children understood certain things about desires before achieving parallel insights about beliefs. For example, as outlined in Table 1, on simple tasks, children understand diverse desires before understanding diverse beliefs. A progression from reasoning about desires to reasoning about beliefs also characterizes children’s conversations (15, 16).

More extended progressions of understanding further characterize theory-of-mind development and have been useful in illuminating the origins and mechanisms of development and change. Consider the theory-of-mind scale (17) that encompasses carefully constructed tasks assessing children’s understanding of all the distinctions in Table 1. Studies using this scale with hundreds of preschoolers in Australia, Canada, Germany, and the United States show a consistent order of difficulty, as seen going from top to bottom in Table 1. Validated progressions like this allow deeper examination of development, including the extent to which theory-of-mind developments are culturally universal or specific, and do or do not depend on specific childhood experiences.
One possibility is that early theory-of-mind achievements represent maturational unfolding of some theory-of-mind device or module. If so, then sequences and timetables for theory-of-mind developments should be universal. Alternatively, perhaps theory-of-mind understandings are the products of social and evidential experiences that vary from child to child and across communities. Tests of such alternatives were undertaken, in part, by comparing Western and Chinese children. Many scholars contrast an Asian focus on people sharing group commonalities and interdependence with a Western focus on people as distinctively individual and independent. Indeed, in conversation with young children, Chinese parents comment frequently on knowing (18), including consensual knowledge that everyone should learn, while American parents comment more on

---

**Figure 1.** Two common false-belief tasks for children.

Note. False-belief tasks have many forms, but two common ones use changed locations (top) or deceptive contents (bottom). Younger children consistently incorrectly say the character will search in the cupboard (where the target object really is) or will think the Band-Aid box has a toy truck inside.

---

**Table 1**

*The Theory-of-Mind Scale Items.*

<table>
<thead>
<tr>
<th>Task</th>
<th>Brief description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Diverse desires</td>
<td>Child judges that two persons (the child versus someone else or two other people) have different desires about the same object (e.g., one likes broccoli and the other does not).</td>
</tr>
<tr>
<td>2. Diverse beliefs</td>
<td>Child judges that two persons have different beliefs about the same object, <em>when the child does not know which belief is true or false</em> (e.g., one thinks an occluded box holds a car and the other thinks it holds a ball).</td>
</tr>
<tr>
<td>3. Knowledge access</td>
<td>Child judges another person’s ignorance about the contents of a container when child knows what is in the container (e.g., child knows drawer hides a toy dog, but child judges that another person who has never seen inside does not know what is there).</td>
</tr>
<tr>
<td>4. Contents false belief*</td>
<td>Child judges another person’s false belief about what is in a distinctive container when child knows what is in the container (e.g., child knows a familiar Band-Aid box has a truck inside, but judges what someone else who has never seen inside will think it contains).</td>
</tr>
<tr>
<td>5. Hidden emotion</td>
<td>Child judges that a person can feel one thing but display a different emotion (e.g., character feels sad but can look happy on his face).</td>
</tr>
</tbody>
</table>

*Other false-belief tasks can be used. For several reasons (17) contents false belief is the task included in the standard five-step scale.*
thinking (15), including differences in thoughts among different individuals.

Accordingly, Chinese preschoolers show a consistent but different theory-of-mind sequence from that of Western children in which knowledge access and diverse beliefs are reversed (19). After early understanding of basic aspects of desire, Western children first appreciate differences in beliefs, whereas Chinese children first appreciate acquisition of and access to knowledge—as do children in Iran (20) and Turkey. The two sequences are crucially different (indicating experience-dependent processes of theory-of-mind learning) yet notably similar (indicating robust, universal theory-of-mind development).

Theory-of-mind timetables differ as well, sometimes dramatically. An early discovery that false-belief understanding is substantially delayed in autistic children led to the theory-of-mind hypothesis for autism (21). But autism, with its neurologically based, across-the-board delays, could have its own delayed maturational timetable. Yet deaf children, who do not suffer from the same central neurological impairments and retardation as children with autism, also often have serious theory-of-mind delays.

About 95% of deaf children are born to hearing parents and (unlike deaf children of deaf parents) grow up with early experiences that are very different from their hearing peers. For example, hearing parents mostly communicate didactically with their young deaf child using simple signs or gestures to refer to objects. Moreover, their deaf children are likely to have restricted play with others, resulting in less access to free-flowing social—communicative interactions, and little exposure to discourse about internal states like thoughts and emotions. Deaf children of hearing parents are substantially delayed in understanding false beliefs, like children with autism (22). More comprehensively, they have consistently delayed sequences of understandings on the theory-of-mind scale, taking 12 or more years to progressively achieve theory-of-mind insights that hearing children achieve in 4–6 years (23).

Infants’ Theory of Mind

Theory-of-mind research began with older children and scrolled backward toward infancy. The earliest examples of psychological construals of people appear in infants’ understandings of intentional action and experience; by the end of the first year, children begin to treat themselves and others as intentional agents that have internal experiences.

Consider infant gaze following. Conceivably, infants may just automatically match an adult’s gaze, without any deeper understanding. In fact, young infants often gaze follow adults wearing blindfolds. Yet, when 12-month-olds have experience with blindfolds occluding their own vision (24), they are significantly less likely to follow a blindfolded adult’s gaze. By 18 months, infants do not often gaze follow a blindfolded adult—presumably they now understand that blindfolds occlude visual experience. But when they have experience with a special blindfold that looks opaque yet is easily seen through, 18-month-olds follow the head turns of adults wearing that blindfold. Thus, by 12–18 months, infants have a sense that people’s visual experiences control their gaze (which represents more than just overt eye or head directedness).

In the last 10 years, researchers have claimed that infants go beyond an understanding of intentional actions and experiences: By 10–15 months, they recognize that people act on the basis of their beliefs and false beliefs. The top of Figure 2 presents an initial influential task by Kristine Onishi and Renee Baillargeon (25). Other demonstrations have accumulated step by step. Although these studies are clever and revealing, how their findings should best be interpreted remains controversial.

For example, such findings are often interpreted in deeply nativist fashions as revealing initial understandings that emerge without learning. But recent research with deaf infants of hearing parents shows that they do not show the same false-belief responses that hearing infants do (26). In my view, experience-dependent (constructivist) learning characterizes theory-of-mind understanding from its beginning. Regardless of whether these studies reveal that infants truly understand false beliefs, they confirm that infants typically understand agents as goal directed, that agents’ changing experiences yield for them awareness or unawareness of key events, and that aware and unaware agents act differently.

One pressing question from this research with infants concerns how infants’ understandings relate to later theory-of-mind accomplishments (which, recall, shape children’s social actions and interactions profoundly). At the least, we now know that individual differences in infants’ social-cognitive understandings longitudinally predict differences in the timing of preschoolers’ theory-of-mind achievements (27, 28).

Chimpanzees and Dogs

What sorts of theory-of-mind achievements are apparent in our closest primate relatives? This provocative question, posed by Premack and Woodruff, has an extended history. For many years, the conclusion was: none (29, 30). Primates performed poorly on all sorts of tasks (many with parallels in research on human infants) designed to demonstrate awareness of the internal states of others. But this early research used cooperative—communicative paradigms, where success depended on appreciating that someone’s goal was to help (e.g., share food with the chimpanzee). In competition paradigms, chimpanzees performed more optimally. For example, take a situation in which a piece of food is placed between a dominant and a subordinate chimpanzee and, because of various visual obstacles, the two have differing awareness of the food; when subordinates can see two pieces of food and the dominant chimpanzee sees only one, subordinates preferentially target the less-risky food that the dominant chimpanzee cannot see. Studies suggest that chimpanzees demonstrate these preferences because they understand something about the link between seeing and knowing (31): They
adjust their behavior not only on the basis of what dominants can currently see, but also on the basis of what dominants have and have not seen in the past. Rhesus monkeys also show impressive sensitivity to others’ perceptual experiences in competitive situations.

Unlike chimpanzees, domestic dogs perform well on simple tasks where they read the cooperative–communicative intentions and experiences of humans. One influential hypothesis is that dogs evolved this human-infant-like social-cognitive prowess in their long history of domestication, with the key being the domestication of their temperaments (to be nonaggressively, nonfearfully attentive to humans; 32). This temperament hypothesis led to research with young children. Children (even infants) who are nonaggressively, nonfearfully attentive to others perform more optimally on theory-of-mind tasks. And they do so concurrently and also in longitudinal research where early measures of observant-reflective social temperament predict theory-of-mind achievements several years later (33).

Developments Beyond Preschool
Children’s understanding of mind and of people develops in important respects beyond the ages of 5 or 6 years. For example, only after the preschool years do children develop a deepening appreciation of the mind as different from the brain. When asked whether they can perform various kinds of functions without a brain, and separately without a mind (34, 35), the youngest children respond identically about the brain and mind, and they conceive of the mind/brain as needed exclusively for purely mental acts. Only by fifth grade or so do children become generally aware that the brain differs from the mind.

For me, one of the most intriguing later developments concerns children’s increasing willingness to entertain ideas of extraordinary minds and capacities. Initially, Justin Barrett (36) demonstrated that as children come to appreciate the constraints of ordinary human knowledge and belief—for example, that people can have false beliefs—they recognize that God could have more extraordinary powers. Many findings have followed from this, charting children’s understanding of omniscience, afterlife, souls, and the like (35, 37). The school-age years are pivotal for children’s understanding of such extraordinary experiences (even for those children in devout homes who receive instruction and exposure to these ideas very early in life; 38). These studies show that extended progressions in children’s theories of mind are built on early preschool understandings that provide the foundation for children’s construction of later ideas, including
their receptivity to and assimilation of sociocultural teachings, doctrines, and ideas about God, superheroes, Santa Claus, and more.

GOING FORWARD

It is easy to think of key issues for further research. More research is needed on how theory of mind operates in adulthood, including late life—researchers are just beginning to consider if theory of mind is party to the general declines of cognitive aging or resistant to decline. Systematic research on how early theory-of-mind differences affect children’s later educational achievements (e.g., in reading, math, and science) has only begun (for a review see 39). Understanding extraordinary minds goes beyond agents like God and superheroes. Consider robots and personified smart technological devices (e.g., Siri, Echo, Alexa); children live in a world that increasingly includes such devices. How do they think and feel about such devices and how does this affect their interactions with and learning from them?

Infants

Despite the boom in research with infants, it is deeply incomplete: Both developmental research and research specifying where infants fail as well as succeed are needed. Too often, demonstrations that infants apparently understand X (false belief, say) do not test or report boundary-setting conditions where the same infants fail.

For example, if infants understand false belief in Baillargeon-type paradigms (see the top of Figure 2), they also must understand that seeing leads to knowing and not seeing leads to ignorance. But a few studies suggest they do not, including the tasks used by Beatte Sodian and Claudia Thoermer (40) at the bottom of Figure 2. In a true-belief condition for that study (not shown in Figure 2), children watched an agent that saw all the movements of a toy that first went into a gray box but then transferred to a white box (paralleling the Onishi and Baillargeon false-belief task, except the agent saw everything). Accordingly, infants then looked longer at the reach-gray test event (because, given a true belief, the agent should search where the toy actually is). Furthermore, in the ignorance condition, infants apparently understood that not seeing leads to ignorance because they did not look longer to either test event (ignorant agents could search anywhere). However, the true-belief-after-delay condition shows they probably do not fully understand that seeing leads to knowledge (or true belief). In that case, although the person saw the ball go into the gray box, infants did not look longer at the white box test event. The agent saw all the relevant movements, and was absent for an irrelevant short time when nothing happened, but infants failed to understand the agent’s true belief. Infant successes occur amid as-yet-unknown failures.

Moreover, studies with infants typically report on 15-month-olds alone, or 18-month-olds alone (and moreover, on different tasks). We need a progressive developmental picture. Research on infants’ social-cognitive learning would also be helpful. In initial studies, infants applied statistical learning to acquire information about social agents and even to infer their mental states (41). How much of infant theory-of-mind development might this account for?

Cognitive Neuroscience

Investigations with adults demonstrate that theory-of-mind reasoning involves a network of neural regions, most consistently the medial prefrontal cortex, and the left and right temporoparietal junction, but also several temporal lobe sites (42). These regions are recruited when adults engage in mental-reasoning tasks, and they are impaired in autistic adults.

Even if findings from studies with adults were crystal clear—and they are not—they could not provide an understanding of brain and cognition earlier in development. Thus, direct neurocognitive examinations of younger children are needed—especially in children from 2 to 6 or 7 years, when developmental changes are pronounced. Such developmental neuroscience is just beginning (43, 44). Emerging research has already begun to show developmental changes in preschoolers in the theory-of-mind network, which would seem unlikely if that network were mature from the start and if theory of mind after infancy merely reflected changes in executive functions or language, as some propose (45). Changes in this network are also emerging in older children, with the potential for more effectively illuminating theory-of-mind changes after the preschool years.

Nonhumans

Research on the theory-of-mind accomplishments (and limits) of chimps and dogs is also not yet very developmental. Mostly, such studies have looked at adult animals. How have their now-mature capacities developed? Perhaps insights that humans acquire easily in life are mostly late-developing insights for chimps and dogs. Developmental research with animals would help us understand more optimally the phylogenesis and ontogenesis of social cognition. More detailed information on other nonhuman species would also be informative. Recent work with birds seems particularly striking (46).

CONCLUSIONS, SO FAR

Over 30 years, the field of theory of mind has emerged, developed, and changed. So have I, although I remain not bored with it all. Instead, I am impressed and energized by how much has been accomplished. We began in the 1980s focusing on preschoolers, but now chart theory-of-mind achievements from infancy through adulthood, from the nursery to the schoolyard to the classroom and into the highways and byways of social life. We began with behaviors and now probe neural networks, genes, and social networks. We began looking at children in a few Western locales and now look worldwide. We began with
nothing but questions; now we have many answers, though of course, answers provoke new questions. So one key accomplishment is a firmer sense of how much remains to be known.

REFERENCES