2.7

The New Puzzle of Theory of Mind Development

REBECCA SAXE

Human children make a remarkable discovery: Other people have minds, similar to but disconnected from their own. Other people see a slightly different world; have different desires, preferences, and values; and have different knowledge or beliefs from their own. That is, other people’s minds contain representations of the world that are often true and reasonable but may be perverse, incomplete, or even totally false. This discovery helps children to make sense of some otherwise mystifying behavior: why mom would eat broccoli even though there is chocolate cake available (e.g., Repacholi & Gopnik, 1997), for example, or why she is looking for the milk in the fridge even though dad just put it on the table (e.g., Wimmer & Perner, 1983). Beyond simple action understanding, though, inferences about what other people know, want, or believe (collectively called a “Theory of Mind”; Gopnik & Wellman, 1992) inform children’s growing understanding of many aspects of human social life: empathy, morality, deception, metaphor, irony, and fiction (e.g., Baird & Astington, 2004; Capelli, Nakagawa, & Madden, 1990; Peterson, Wellman, & Liu, 2005; Winner & Leekam, 1991). Thinking about other minds becomes one of the most impressive, and distinctive, accomplishments of human abstract cognition.

How, and when, do children learn to think about other people’s thoughts? In developmental psychology, most research has focused on one key transition in this developmental process: when children understand that people can have false beliefs. In the past three decades, thousands of children, in hundreds of studies around the world, have been shown a scenario involving a simple false belief: for example, mother thinks the milk is in the fridge, but really it is on the table. Children are asked: “Where will she look for the milk?” or “Why is she looking in the fridge?” If you’ve never tried asking a 3-year-old this kind of question, I strongly encourage it. It is astonishing to watch a bright, articulate, verbal child confidently predict that she will look for the milk on the table, and if she is looking in the fridge, she must not want the milk. Five-year-old children, by contrast, usually predict that she will look in the fridge, because that’s where she thinks the milk is (Wellman, Cross, & Watson, 2001).

This profile of developmental change in children’s Theory of Mind is so reliable, across methods and across cultures (e.g., Avis & Harris, 1991; Liu, Wellman, Tardif, & Sabbagh, 2008; Wellman, Cross, & Watson, 2001), that it has become diagnostic of typical human social development. Children with autism spectrum disorders, for example, are specifically delayed in understanding false beliefs (e.g., Baron-Cohen, Leslie, & Frith, 1985; Leslie & Thaiss, 1992; Peterson, Wellman, & Liu, 2005). And nonhuman animals, in spite of rich social cognition in other respects, never quite reach a full and flexible understanding of false beliefs (e.g., Kaminski, Call, & Tomasello, 2008; Santos, Flombaum, & Phillips, in press).

Something about the typically developing human brain seems to be specially designed to help children make this critical cognitive leap. The advent of neuroimaging allowed researchers to look directly into human brains to find it—in what may be the most spectacular novel contribution of functional neuroimaging to cognitive science. In human adults, a group of brain regions is specifically devoted to social cognition (e.g., Gallagher et al., 2000). At least one of these, located near the right temporo-parietal junction (and therefore frequently called “the RTPJ” for short), is active specifically when people are thinking about other people’s thoughts (Saxe, in press).

This introduction takes us to the state of the art around 2005, when developmental psychology
and cognitive neuroscience appeared to provide converging support for a strong theory of when and how children acquire a Theory of Mind. Building on simpler foundations (e.g., concepts of "agent," "action," "intention," and "perception"); Csibra & Gergely, 2007; Woodward, 2009), children make a key leap in their understanding of other minds between ages 3 to 5 years. Adult Theory of Mind depends distinctively on a group of brain regions, predominantly including the RTPJ. So it seemed plausible that key maturational changes typically occurred in the RTPJ between ages 3 and 5 years, supporting the cognitive advances, and that this maturation was specifically and adversely impacted by the etiology of autism (e.g., Saxe, Carey, & Kanwisher, 2004). Indeed, this view still seems appealing to me.

Recently, though, aspects of this picture have begun to unravel, creating a new puzzle of Theory of Mind, a key challenge for our understanding of social development. The puzzle is as follows: Recent advances in developmental psychology suggest that children have some understanding of false beliefs much earlier than age 3 years, and initial neuroimaging studies of children's brains suggest that key maturational changes in the RTPJ occur much later than age 5 years. To accommodate these data, a new theoretical picture will be needed.

Most of the troublesome new developmental data come from studies of infants' gaze. Before infants can talk or follow instructions, their eye movements already reveal their expectations about the world around them. Infants look longer at an object or event that surprises them and will look anticipatorily at a place where they expect something interesting to happen. Capitalizing on these simple behaviors, researchers have designed experiments that ask infants what they expect to happen when a person has a false belief. If mom last saw the milk in the fridge, even if it is now at the table, will infants look anticipatorily toward the fridge, expecting her to open it? If she goes instead to the table, will infants treat that action as surprising, and look longer than if she had gone to the fridge? In essence, the recent experiments answer: yes (e.g., Onishi & Baillargeon, 2005). Young infants appear to expect people to act in accordance with their beliefs about objects, whether they are true or false.

How young are these infants? Slightly confusingly, it differs across experiments. Some find evidence for understanding of false beliefs in 24- but not 18-month-olds (Southgate, Senju, & Csibra, 2007), whereas others find it in 18-, 15-, or even 13-month-olds (Scott & Baillargeon, 2009; Song & Baillargeon, 2008; Surian, Caldi, & Sperber, 2007). In all of the studies, though, the children are much younger than 3 years old—the age of the classic, reliable, and quite dramatic failures on traditional measures of false-belief understanding.

So, how can infants successfully make the very inferences that elude 3-year-old children? It is tempting to answer that the old studies were wrong and underestimated the 3-year-olds. Perhaps the experiments were unnecessarily complicated, and/or the young children were confused, rushed, and intimidated, or trying to give the answer the experimenter wanted. If the experimenters would just proceed more slowly and clearly, and give the children more practice, 3-year-olds (and even 2-year-olds) would show their true competence. The problem with this answer is that it has been tested and proven wrong. Hundreds of attempts to make the experiment simpler and the question clearer (e.g., Where will mom look first for the milk?) made only a modest difference at best (Wellman, Cross, & Watson, 2001), and 3-year-olds still predict that mom will look in the fridge even after dozens of trials of practice over many weeks (Baker & Leslie, 2008).

Renee Baillargeon, who pioneered the study of false-belief understanding in infants, argues that predicting where mom will look requires children to make inferences about the invisible future; in this demanding context, children go for the salient easy answer (where the milk actually is) because they do not have the cognitive resources left over to inhibit the easy answer and retrieve the more complicated answer (where she thinks it is), which they also know (Baillargeon, Scott, & He, 2010). Again, though, experimental data disagree. In some experiments, 3-year-olds get to watch mom look in the fridge and are then asked why she did that; there is no future to predict, and the "easy" answer should now be the one right in front of their eyes. Still 3-year-olds do not just say what infants appear to know, that mom thinks the milk is there. Instead, they generate whole new explanations (Goodman et al., 2006; Moses & Flavell, 1990): Apparently she does not want milk, so maybe she is looking in the fridge because she wants orange juice.

Another way out of the bind is to claim there is something wrong with the way the infant studies are conducted, so that the infants' concepts are being overestimated. Perhaps something else is catching infants' attention in the scenarios they are watching, creating differences in gaze without any real understanding of false belief. But this is
also too easy. Specific evidence for infants’ understanding of false beliefs now comes from almost a dozen studies done in multiple labs, across multiple countries (Scott & Baillargeon, 2009; Song, Onishi, Baillargeon, & Fisher, 2008; Southgate, Senju, & Csibra, 2007; Surian et al., 2007). More generally, in other domains of cognition, measurements of infants’ conceptual development based on gaze behavior converge perfectly with more traditional measurements like reaching actions (e.g., Feigenson & Carey, 2005; Sommerville, Woodward, & Needham, 2005). It will not be easy to simply write off the gaze measurements as meaningless or irrelevant.

Instead, it seems that both sets of data are right: Infants do expect people to act in accordance with false beliefs, but 3-year-olds do not predict or explain actions in terms of false beliefs. In fact, individual children can show both patterns simultaneously, looking anticipatorily at the fridge, but predicting verbally (and betting very confidently) that mom will go to the table (Ruffman, Garnham, Import, & Connolly, 2001). At this point, many developmental psychologists, myself included, are drawn to the idea that there must be two different ways of understanding false beliefs. Young infants have one way; it is simple, fast, and efficient, but it is limited in scope and precision. So infants can formulate accurate expectations about others’ actions in some simple contexts, like reaching for the milk you saw a moment ago. This system might be called an “implicit” Theory of Mind; it supports action understanding in the moment, but not reflection, deliberation, or revision. This simple system cannot expand to handle the complexities of adult Theory of Mind, though, so young children have to develop a whole second system of concepts of beliefs, desires, and actions, which has much larger scope and greater precision, is more flexible and easier to expand: an “explicit” Theory of Mind.

The proposed distinction between implicit and explicit Theory of Mind gains plausibility because a similar pattern occurs in another area of cognitive development: infants’ understanding of numbers (e.g., Apperly & Butterfill, 2009). Very young infants do understand something about number and quantity, but their understanding is limited in scope and precision. As they slowly acquire their culture’s words and concepts for numbers, they can acquire massively expanded computational power and flexibility (e.g., Le Corre & Carey, 2007).

Unfortunately, there is no actual evidence supporting a distinction between implicit and explicit Theory of Mind. So far, no one has found limits in scope or precision of infants’ Theory of Mind. Nor is it perfectly clear why eye movements should reveal infants’ implicit understanding, while pointing or talking requires the later developing explicit theory, even when giving the same answer to the same question. By contrast, the limits on young infants’ numerical concepts are evident both in eye movements and reaching actions (Feigenson & Carey, 2005).

One of my chief hopes is that measurements of brain activity will help resolve this puzzle. If infants and children both have the same Theory of Mind, but different abilities to use inhibition to answer difficult questions, we should see it in the pattern of their brain activity. In adult brains, thinking about thoughts engages one group of brain regions, while solving difficult problems with high inhibitory demands depends on a distinct group of brain regions. If Baillargeon’s theory is right, older children should show different brain activity from younger children (correlated with false-belief predictions) in the “problem-solving” brain regions, but not in the Theory of Mind brain regions. By contrast, if infants and children understand false beliefs using two different mechanisms—one “implicit” and the other “explicit,” for example—we predict a different neural pattern. Infants and 3-year-olds watching mom looking in the fridge should show activity in different brain regions from the classic Theory of Mind pattern, while 5-year-olds’ brains should look like adults’.

It is just now becoming possible to test these questions, by using functional neuroimaging of young children’s brains. For better and for worse, though, the few neuroimaging studies of Theory of Mind development only deepen the puzzle. The contrasting hypotheses derived from developmental psychology predict change in Theory of Mind brain regions around age 1 year versus 4 years; the neuroimaging data show changes in Theory of Mind brain regions around age 9 years (Gweon, Dodell-Feder, Bedny, & Saxe, 2012; Kobayashi, Gloverb, & Templec, 2007; Saxe, Whitfield-Gabrieli, Scholz, & Pelphrey, 2009).

The pattern of functional change we have observed is fascinating. In adults, the RTPJ is highly specialized for thinking about thoughts; you use your RTPJ to think about someone’s beliefs and desires, but not to think about their appearance, social background, or personality traits (Saxe & Kanwisher, 2003; Saxe & Powell, 2006). In 5- to 8-year-old children, the RTPJ shows robust activity when thinking about anything about another person: their thoughts and plans but also their appearance, relationships,
and social background. In late childhood, then, this brain region seems to change its function by becoming more specialized (Gweon et al., unpublished data; Saxe et al., 2009). Starting with a more general role in social cognition, the RTPJ gradually focuses on one specific cognitive challenge: thinking about thought. It would have been very gratifying and simple, if this pattern of functional maturation in the brain underlay children’s growing ability to understand false beliefs around age 4 years. But it does not.

Of course, the neural measurements do not directly contradict either of the developmental hypotheses. There are only a handful of neuroimaging studies of children’s Theory of Mind. The measurement tool most commonly used in adults, functional magnetic resonance imaging (fMRI), is challenging for young children, since it requires lying completely still in a small, loud tube for at least half an hour. To date, no fMRI studies of Theory of Mind have been conducted with children younger than 5 years old. Key functional changes may very well occur in younger children’s brains, but they have not yet been observed. Indeed, using another more child-friendly neuroimaging technique, electroencephalograms (EEG), Mark Sabbagh has found evidence of biological maturation in RTPJ, specifically related to 4-year-olds’ action predictions based on false beliefs (Sabbagh, Bowman, Evraire, & Ito, 2009).

Here is one possible resolution. Young infants’ foundational understanding of other people is rich enough to support some basic understanding of false belief, but it is in other respects seriously limited in scope, precision, and flexibility; in ways we will discover. Perhaps infants can only attribute beliefs and desires about objects and events that are currently present in the environment, or only about simple features (like location and identity), or only with very simple logical structures (a simple proposition). Correctly designed tasks will reveal these limits, in both eye gaze and other behavioral measurements. Adults also use the same fast efficient system in certain contexts; and so the limits on this system will appear on appropriate tests of spontaneous Theory of Mind in adults. Moreover, infants’ Theory of Mind is housed in a distinct neural system, not in the RTPJ; and this system is functionally mature early in development. This alternative neural system will be revealed by future neuroimaging studies of infants.

Then, beginning around age 2 years but continuing throughout childhood into adolescence, children slowly form a second system of concepts for understanding other minds. These explicit concepts are much more computationally powerful and flexible, allowing us to attribute beliefs and desires on any topic, and with much more complex logical structure. They are formulated as common explanations of multiple kinds of evidence: children’s own experience of their own mind, their observations of human behavior, the intuitions derived from their foundational (culturally universal) concepts of actions and agent, and the (culturally specific) verbal descriptions of the mind provided by people around them. The later system is housed in the slowly developing RTPJ (among other brain regions), and it provides the basis for explaining one’s own and others’ behaviors in a wide range of contexts, for moral judgment, for strategic planning, for spreading gossip, and for writing fiction.

I am drawn to this picture. So far there is very little evidence to support or contradict it, but many of the predictions are testable with behavioral and neural measurements. I worry, though, that it is an evasion; whenever cognitive scientists encounter two sets of data that do not fit nicely into one of our existing theories, we simply posit two separate cognitive systems. It would be much more satisfying to construct a single theory, in which each of the observed patterns arises as a special case or in a defined context (e.g., Saxe, 2005; Tenenbaum & Griffiths, 2001). But I do not yet know what such a theory would be.

I do believe that the new puzzle of Theory of Mind is one of the key open challenges in the science of social development, and in cognitive science more broadly. Once again, we need a new theory of Theory of Mind.

REFERENCES


The New Puzzle of Theory of Mind Development


2.8

How Real Is the Imaginary?

The Capacity for High-Risk Children to Gain Comfort From Imaginary Relationships

MARJORIE TAYLOR AND NAOMI R. AGUIAR

He kind of looks like me . . . we like to play sword fighting . . . we pretend to play Lego Star Wars. He's super funny . . . nice, generous, crazy sometimes. He tells me jokes . . . shares his snacks with me . . . When I'm upset, he makes me talk sometimes; he makes me feel a little bit better. He's really good at doing things . . . helping me with my homework. He can run super fast. He's super nice . . . a good friend.

—9-year-old boy's description of his friend "Bob"

This description includes many of the features that signal high quality in a friendship—a combination of recreation, guidance, validation, and intimate exchange (Parker & Asher, 1993). The support and companionship of such friends are important to our happiness and psychological health throughout life and in childhood are widely believed to promote social competence and resilience. Thus, having a friend like Bob might seem to be obviously a good thing—except for a caveat. Does it matter that Bob is imaginary? That he eats bugs, can read a 150-page book in a minute, and likes to shape shift into animals or sometimes plants (e.g., a tree)?

Imaginary companions are invented characters that children interact with and/or talk about on a regular basis. (Sometimes they are invisible and sometimes they are based on special stuffed animals or dolls referred to as personified objects.)

There is a history of ambivalence about the meaning and significance of such friends. On the one hand, Developmental Psychology has backed away from Piaget's view that they reflect "a lack of coherence . . . an immature mind that had not yet adapted to reality" (1962, p. 131), and recent studies have overturned the lay stereotype of the child with an imaginary friend as a shy, unusual, and withdrawn individual with emotional problems. The findings from our lab and elsewhere show that in Western cultures having an imaginary companion during the preschool years is relatively common (Singer & Singer, 1990; Taylor, 1999) and tends to be associated with positive characteristics such as referential communication skills (Roby & Kidd, 2008) and narrative depth in storytelling (Trionfi & Reese, 2009). And contrary to the stereotype, imaginary companions do not seem to be associated with fewer friendships with real children (Gleason, 2004). Yet imaginary companions are quick to elicit adult concern and suspicion if they are not well behaved—even though many children describe their imaginary companions as unruly, bossy, or disobedient (Taylor, Carlson, & Shawber, 2007)—or if they appear after the preschool years—even though it is not particularly unusual for older children and adolescents to have them (Pearson et al., 2001; Seiffge-Krenke, 1997).

While not all friendships, be they real or imaginary, have the depth and stability to become significant in our lives, do at least some imaginary companions evolve into an important source of real-world support? This is what we mean when we ask how "real" is the imaginary. Note that we are not asking if children are confused and think their imaginary companions are real. For the record, the evidence clearly indicates that even preschool children are well aware that they are pretend (Taylor & Mottweiler, 2008; Taylor, Shawber, & Mannering, 2009). Children who participate in our research frequently make explicit and spontaneous references to the fantasy status of their imaginary companions (e.g., "I just