CHAPTER

20

Developmental Cognitive Neuroscience of Theory of Mind

H. Gweon, R. Saxe

Massachusetts Institute of Technology, Cambridge, MA, USA

OUTLINE

It's the best possible time to be alive, when almost everything you thought you knew is wrong.

***Arcadia*, by Tom Stoppard**

Imagine you arrive back at the laundromat and see a stranger take your clothes out of the dryer and start to fold them. What is going on? With only seconds of perceptual data about this stranger, you can immediately conjure up multiple plausible explanations: maybe he intends to steal your clothes, maybe he is feeling amazingly generous and wants to help someone out, or maybe he just falsely believes that those are his own clothes. In this example and in countless other brief and extended social interactions every day, we do not just describe people's actions as movements through space and time. Instead we seek to explain and judge and predict their actions, and we do so by appealing to a rich but invisible causal structure of thoughts, beliefs, desires, emotions, and intentions inside their heads. This capacity to reason about people's actions in terms of their mental states is called a 'theory of mind' (ToM).

This chapter is about what we know, and what we do not know, about how the human brain acquires its amazing capacity for ToM. In the past few decades, ToM has been studied intensively in childhood development (using behavioral measures) and in the adult human brain (using functional neuroimaging). Converging evidence from these two approaches provides insight into the cognitive and neural basis of this key human cognitive capacity. However, as we highlight later, we are especially excited about the future of ToM in developmental cognitive neuroscience: studies that combine both methods, using neuroimaging methods to directly study cognitive and neural development in childhood.

We start by describing an account of ToM, in development and in neuroscience, that we shall call the 'Standard' view. Next, we describe some recent challenges that shake the foundations of the Standard view. Finally, we point to the open questions, and especially the key contributions that developmental cognitive neuroscience can make in the next generation of studies of ToM.

20.1 WHAT WE THOUGHT WE KNEW: THE STANDARD VIEW

20.1.1 Development

The laundromat example makes clear a central feature of ToM: it is especially useful when other people have false beliefs. When strangers fold their own laundry,

Neural Circuit Development and Function in the Brain: Comprehensive Developmental Neuroscience, Volume 3 http://dx.doi.org/10.1016/B978-0-12-397267-5.00057-1

no special explanation seems warranted. We can describe their actions in behavioral terms: folding laundry is a thing people frequently do in laundromats. However, when a stranger starts to fold your laundry, then it becomes important to figure out: what are they thinking? The understanding that they may have a false belief makes an otherwise highly unlikely action suddenly predictable. If they believe it is their own laundry, then of course they will start folding it.

Because false belief scenarios are so diagnostic of ToM inferences, understanding of false beliefs has been considered a key milestone in ToM development. Children's ability to predict and explain actions based on a false belief is typically assessed in a 'false belief task' (see Figure 20.1). In a typical example, children hear a story like this one: Sally sees her dog run to hide behind the sofa; then, while Sally is out of the room, the dog moves to behind the TV. Children are then asked to predict where Sally will look first for her dog when she returns to the room (Baron-Cohen et al., 1985; Wimmer and Perner, 1983).

Adults immediately recognize that Sally will look for her dog behind the sofa, where she thinks it is. Surprisingly, 3-year-olds systematically make the opposite prediction; they confidently insist that Sally will look behind the TV, where the dog really is (Wellman et al., 2001). Moreover, if 3-year-olds actually see Sally looking behind the sofa, they still do not appeal to Sally's false belief to explain her action, but instead appeal to changed desires (e.g., 'she must not want the dog,' Goodman et al., 2006; Moses and Flavell, 1990). In contrast, typical 5-year-old children correctly predict and explain Sally's action, by appealing to her false belief.

This pattern of children's judgments, over development, is incredibly robust; it has been replicated in hundreds of studies conducted over four decades. The same shift in understanding false beliefs and the ability to use beliefs to explain actions occurs between 3 and 5 years in children from rural and urban societies, in Peru, India, Samoa, Thailand, and Canada (Callaghan et al., 2005) and even in children of a group of hunter-gatherers in Cameroon (Avis and Harris, 1991).

Developmental psychologists do not disagree about these data; they disagree about the interpretation. To start with, we will articulate two claims we call the 'Standard' interpretation of these results. This is the view that most informed, and converged with, the first neuroimaging studies of ToM in the adult brain. It was never, however, a consensus opinion; in whole and in part, every aspect of the Standard view has been hotly debated all along.

First, the Standard view of development on the false belief task proposes that children undergo a key conceptual change in their ToM between ages 3 and 5 years. They

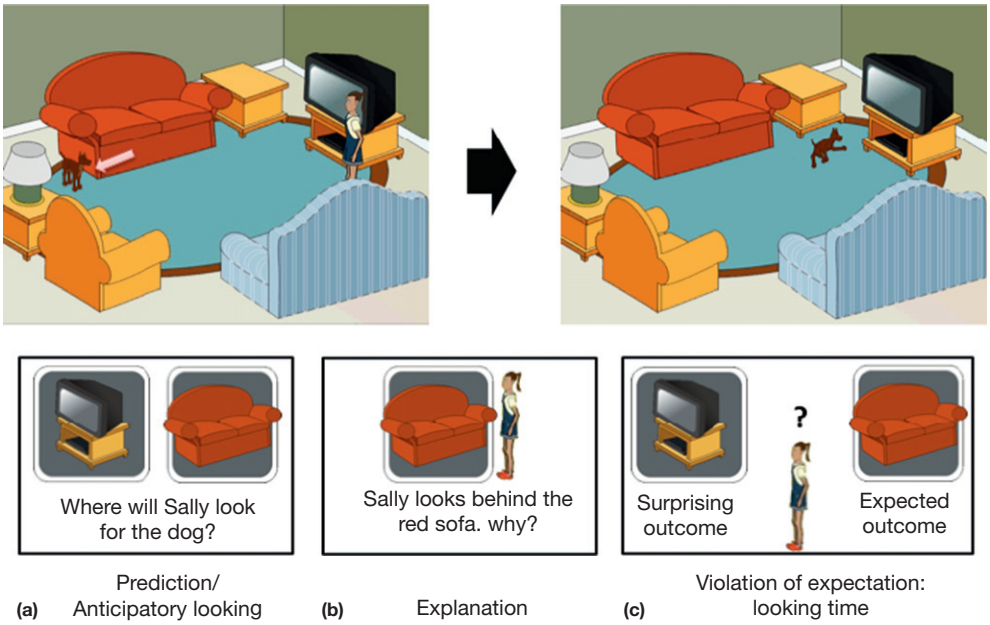

FIGURE 20.1 An example of a standard false belief scenario. Sally sees her dog hide behind the red sofa, but the dog moves to behind the TV while she is out of the room. Children's understanding of false beliefs at different developmental stages can be assessed using various methods: (a) by asking children to predict where Sally will look, or measuring their anticipatory look when Sally comes back to find her dog; (b) by asking children to explain Sally's action; or (c) by comparing infants' looking times to observing Sally go toward the sofa (the 'expected' outcome, if they understand false beliefs) or the TV ('unexpected' outcome). In spite of the logical similarity between these tasks, children make the correct prediction/explanation (Sally will look behind the sofa) around age 4 years, but make correct anticipatory looks/longer looking to the unexpected outcome around age 15–24 months. Art thanks to Steven Green.

are slowly acquiring a fully 'metarepresentational' ToM (Perner, 1991), which lets them understand people's beliefs and thoughts as representations of the world. Representations are designed to accurately reflect the real world, but sometimes fail to do so. So a representational ToM allows children to understand when and how the content of a person's belief can be false (Gopnik and Astington, 1988; Wellman et al., 2001), and that in these cases, people's actions will depend on what they believe, not on what's really true.

Note that learning to understand false beliefs involves change within a child's ToM, not the acquisition of a ToM. Even very young infants understand that people's actions depend on what they want (e.g., Phillips and Wellman, 2005; Woodward, 1998; see Gergely and Csibra, 2003 for a review), and what they can see (Meltzoff and Broks, 2008). That is already a sophisticated ToM, and it makes the right predictions for people's actions in many, if not most, circumstances. Specifically, when someone has a false belief, though, a ToM based mainly on understanding intentions makes the wrong prediction: if Sally wants her dog, she will go get her dog, so she will go where it is, behind the TV. Thus, on the Standard view, young children do have a ToM founded upon very early developing concepts of intention and perception. Nevertheless, their ToM changes substantially between ages 3 and 5 years by the addition of a full concept of 'belief.'

Second, the Standard view proposes that changes in false belief understanding reflect maturation of a 'domain-specific' mechanism for ToM. Between the ages of 3 and 5 years, children change and mature in many ways: they come to have a richer vocabulary and a better memory and a larger shoe size. However, on the Standard view, ToM develops separately: ToM task performance is not just a matter of getting smarter or faster in general, but specifically of conceptual change within ToM.

One way to assess domain specificity is to compare children's development of reasoning about false beliefs with their ability to solve very similar puzzles about other false representations: outdated photographs. For example, children might see Sally take a photograph of the dog behind the sofa; after the dog moves to the TV, the children are asked where the dog is in the photograph. The false photograph task is logically very similar to the false belief task, requiring very similar capacities for language, memory, and inhibitory control (e.g., the ability to choose between two competing response alternatives). Nevertheless, young children are significantly better at the false belief task (Zaitchik, 1990), possibly because they have a 'special mechanism' for ToM which gives their performance a boost.

Stronger evidence that ToM is separate from other parts of cognition comes from studies of children with neurodevelopmental disorders. Children diagnosed with autism spectrum disorders (ASDs) are significantly delayed in passing false belief tasks, compared with typically developing children or children with other developmental disorders like Down syndrome (Baron-Cohen, 1997; Baron-Cohen et al., 1985). On a larger set of tasks, tapping multiple different aspects of ToM (e.g., understanding desires, beliefs, knowledge, and emotions), children with ASD show both delayed development and also disorganized development. That is, typically developing children pass these tasks in a stable order (e.g., understanding false beliefs is easier than understanding false emotions), but children with autism pass the tasks in a scrambled order, as if they were passing for different reasons (Peterson et al., 2005). Furthermore, the difficulty seems to be specific to ToM: compared with typically developing children matched for IQ and verbal abilities, children with autism show comparable (or even better) performance on nonsocial tasks that require similar logical and executive capacities, like the false photograph task (Charman and Baron-Cohen, 1992; Leekam and Perner, 1991; Leslie and Thaiss, 1992).

The observation that a neurobiological developmental disorder, ASD, could disproportionately affect development of ToM supports the Standard view that ToM development depends on a distinct neural mechanism. That is, some brain region, chemical, or pattern of connectivity might be specifically necessary for ToM, and disproportionately targeted by the mechanism of ASD. This hypothesis was difficult to test, though, until the advent of neuroimaging allowed researchers to investigate the brain regions underlying high-level cognitive functions like ToM.

20.1.2 Neuroimaging

The Standard view proposes that children undergo a conceptual change in their ToM between the ages of 3–5 years, from a conception involving actions and goals to one involving beliefs, and that this development of ToM is supported by a domain-specific mechanism. Early neuroimaging studies of ToM in adults appeared to converge nicely with both of these predictions.

Following the tradition in developmental psychology, the early neuroimaging studies of ToM required participants to attribute false beliefs to characters in stories or cartoons. Meanwhile, the scientists measured the oxygen in the blood of the participant's brain, either using radioactive labels (positron emission tomography, PET; Happe et al., 1996) or by measuring intrinsic differences in the magnetic response of oxygenated blood (functional magnetic resonance imaging, fMRI; Brunet et al., 2000; Fletcher et al., 1995; Gallagher et al., 2000; Goel et al., 1995; Saxe and Kanwisher, 2003; Vogeley et al., 2001). At that point, something remarkable happened.

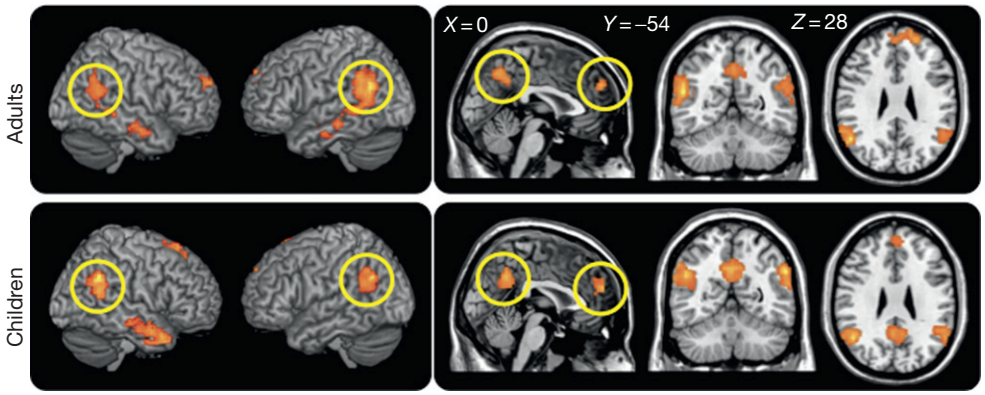

FIGURE 20.2 ToM brain regions in adults (top) and children aged 5–12 years (bottom): the right temporoparietal junction (RTPJ), left temporoparietal junction (LTPJ), precuneus, and dorsal medial prefrontal cortex (yellow circle, from left to right). The bilateral TPJ and precuneus are shown on the coronal section, and all four regions (except for the precuneus in adults) are shown on the horizontal section. *Data shown are from Gweon et al. (2012).*

Across different labs, countries, tasks, stimuli, and scanners, every lab that asked 'which brain regions are involved in ToM?' got basically the same answer: a group of brain regions in the right and left temporoparietal junction (TPJ), right anterior superior temporal sulcus (STS) and temporal pole, the medial precuneus and posterior cingulate (PC), and the medial prefrontal cortex (MPFC; Castelli et al., 2000; Fletcher et al., 1995; Gallagher et al., 2000; German et al., 2004; Goel et al., 1995; Saxe and Kanwisher, 2003; Vogeley et al., 2001; see Figure 20.2). These studies provided initial evidence for a distinct mechanism for ToM, which converged with the predictions of the Standard view.

First, these regions did not respond to other tasks that require similar logical and executive capacities as the false belief task. As described earlier, the 'false photograph' task requires participants to answer questions based on a physical, tangible representation of the past (i.e., a photograph) that used to be true but is currently false. Similar to the false belief task is the false sign task (Parkin, 1994). Understanding a false directional sign involves the use of a symbolic representation that misrepresents the current reality: for example, a signpost indicates that an object is in location A, but the object is then moved to location B. The false photograph and false sign task provide a good test of 'domain specificity' for brain regions; these tasks are very similar to the false belief task in most cognitive demands, and differ mainly in whether they require reasoning about a belief. Using that logic, we can claim that at least some brain regions in human adults are specific for ToM. The bilateral TPJ, MPFC, and PC all respond much more during false belief compared with false photograph stories (Saxe and Kanwisher, 2003). Of these regions, the right TPJ (RTPJ) in particular responds more during stories about false beliefs compared with very closely matched stories about false signs (Perner et al., 2006). Also, identical nonverbal cartoon stimuli elicit RTPJ activity when participants construe the cartoons in terms of a character's false beliefs, but not when participants produce the exact same responses using a nonsocial 'algorithm' (Saxe et al., 2006). So at least the RTPJ, and possibly the other regions in this group, are plausible candidates for the domain-specific mechanism predicted by the Standard view.

Second, brain regions for thinking about beliefs and desires are near, but distinct from, brain regions involved in understanding actions and goals. In the right temporal lobe are brain regions involved in perceiving human bodies and body postures (extrastriate body area, EBA; Downing et al., 2001), movements (MT/V5; Grossman et al., 2000; Tootell et al., 1995), and in particular, people's facial expressions and bodily movements (right posterior STS, pSTS; Howard et al., 1996; see Chapter 19). Interestingly, the activity in the right pSTS depends not just on the action itself; it responds more to actions that are unexpected or incongruent in context (Brass et al., 2007; Pelphrey et al., 2003, 2004). For example, Brass et al. (2007) showed that the pSTS response is enhanced when participants look at a man using his knee to push an elevator button with nothing in his hands, compared with a man performing the same action but with his arms full of books (i.e., when using his knee is rational, in context).

The right pSTS response to intentional actions is impressive, but it would not support your inference about the stranger folding your clothes in the laundromat: in addition to detecting his action as intentional, you specifically need to infer his beliefs and desires (does he know those are your clothes? does he want to help or harm you?) in order to explain why he is doing so. These inferences appear to depend especially on the RTPJ.

The RTPJ is adjacent to the right pSTS (Gobbini et al., 2007), but has a different response profile (see

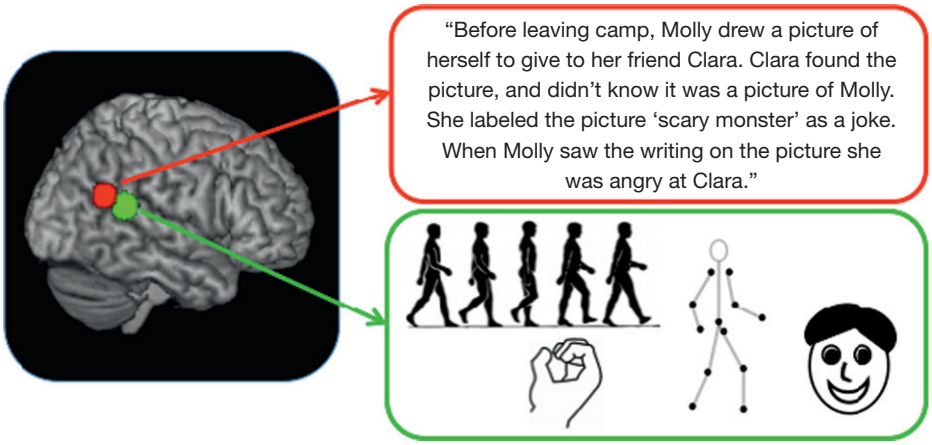

FIGURE 20.3 The RTPJ (red) and the right posterior superior temporal sulcus (pSTS) (green). The RTPJ responds to stimuli that invoke reasoning about thoughts and beliefs, whereas the right pSTS responds to human actions such as walking, grasping, eye gazes, and mouth movements.

Figure 20.3). The RTPJ is recruited during stories or cartoons that depict a person's thoughts, but not in general for depictions of human actions, whereas the right pSTS is recruited when watching movies of simple actions, like reaching for a cup, but not for verbal descriptions of actions. The RTPJ also does not respond to photographs of people (Saxe and Kanwisher, 2003) or to descriptions of people's physical appearance (Saxe and Kanwisher, 2003; Saxe and Powell, 2006). Activity in the RTPJ is low during descriptions of people's physical sensations like hunger, thirst, or tiredness (Bedny et al., 2009; Saxe and Powell, 2006). Even within a single story about a person, the timing of the response in the RTPJ is predicted by the timing of sentences describing the character's thoughts. The response in the RTPJ shows a peak just at the time someone's thoughts are described (Saxe & Wexler, 2005; Saxe et al., 2009).

In sum, neuroimaging studies on ToM with adults provide compelling evidence for a domain-specific mechanism for reasoning about thoughts. Moreover, these studies show that the brain regions involved in ToM are distinct from regions that respond to the perception of goal-directed motion. These results converge nicely with the Standard view, which proposes that children gradually shift from an earlier understanding of goals and intentions (perhaps relying on the pSTS) to the concept of 'beliefs' as reasons for people's actions (relying on the RTPJ).

20.1.3 Developmental Cognitive Neuroscience

The Standard view thus makes a set of testable predictions for the future of ToM studies in developmental cognitive neuroscience (Saxe et al., 2004a).

First, there should be qualitative changes in the anatomy of ToM brain regions, especially including the TPJ and dorsal medial prefrontal cortex (DMPFC), around age 4 years, correlated with children's acquisition of the concept of false belief.

Second, prior to age 4 years, children should rely on early-maturing neural mechanisms for social cognition, especially including the pSTS representation of action, that support recognition of goal-directed actions, but not a theory of mental states like beliefs.

Third, after age 4 years, continued development is more likely to occur outside of the ToM regions, in brain systems that support difficult task performance more generally. That is, children might get generally faster or better at resolving conflicting representations, or more sophisticated in their use of language to describe mental states. But the most momentous shift in the neural mechanisms for ToM would already be complete.

The first prediction of the Standard view is that around age 4, the emergence of children's concept of 'false belief' would be supported by qualitative maturational changes in the ToM brain regions. And, in an initial test, this prediction received impressive support. Sabbagh et al. (2009) first tested a large group of 4-year-old children on standard false belief tasks. As predicted, false belief task performance in this group of children was liminal: some reliably passed the false belief tasks, some reliably failed, and some were intermediate. Next, Sabbagh and colleagues used electroencephalograms (EEG) to measure the amplitude and coherence of alpha waves in the same children's brains, while they were just sitting quietly at rest. These measures are thought to reflect anatomical maturation in a cortical region (Thatcher, 1992). Through an analysis technique called standardized low-resolution brain electromagnetic tomography (sLORETA; Pascual-Marqui et al.,

2002), researchers can use these measures to estimate the current density of alpha waves independently in every region. So Sabbagh and colleagues could then ask: where in the child's brain is maturational change specifically predictive of performance on false belief tasks (controlling for both age, and performance on other demanding tasks)? That is, in which brain regions is the density of the alpha signal best correlated with children's ToM development? The answer was: in the RTPJ and DMPFC – the same two regions most commonly associated with ToM in functional neuroimaging studies of adults!

Sabbagh et al.'s (2009) results are exciting because (1) they offer converging evidence implicating the RTPJ and DMPFC in ToM development, using a completely different method and experimental design, and (2) they support the a priori prediction of an association between anatomical maturation in these brain regions, and performance on standard false belief tasks, when children are around 4 years old.

While source localization methods such as LORETA are shown to render reliable results in estimating the sources of EEG signals (Pascual-Marqui et al., 2002; Wagner et al., 2004), EEG methods still offer much lower spatial resolution compared with other neuroimaging techniques such as functional magnetic resonance imaging (fMRI). So the three basic predictions of the Standard view remain to be thoroughly tested.

But what's most exciting is that the most recent research suggests that each of these predictions is at least partly wrong.

20.2 EVERYTHING WE THOUGHT WE KNEW WAS WRONG

20.2.1 Infants

The biggest challenge for the Standard view comes from a rapidly growing body of research showing that even 15–18-month-old infants understand that people can have, and act on, false beliefs. If so, there is a big lacuna in the foundation of the Standard view. Four-year-old children cannot be acquiring a concept of 'false belief' if that concept is already available to one-and-a-half year olds!

The first report that infants understand false beliefs was a looking-time study by Onishi and Baillargeon (2005). The basic logic of a looking-time study with infants is that first one shows the infants a partial event, setting up an expectation of what will happen next. Then, one shows the infants two possible 'completions' for that event. One of the completions is designed to fit the infants' expectations, and the other completion is designed to violate those expectations. Even preverbal infants can then show which completion they 'expected,' by looking longer at the unexpected completion. Onishi and Baillargeon (2005) made very elegant use of this logic to test what infants expect a person to do when she is acting based on a false belief (see Figure 20.1).

In the original study, 15-month-olds watched an experimenter repeatedly hide, and then retrieve, a toy in one of two boxes (a yellow box and a green box). Then, on the critical trial, the experimenter hid the toy in the yellow box. After a short pause, the toy then moved by itself to the green box. The key manipulation was that either the experimenter watched the toy move ('true belief') or the experimenter turned away, and did not see the toy move ('false belief'). Finally, the experimenter reached either into the green box (where the toy really was) or into the yellow box (where she last put the toy). Which reaching action did the infants expect? Consistent with previous evidence that infants understand goal-directed actions, when the experimenter had a true belief, infants looked longer when she reached into the yellow box, than when she reached into the green box, where the toy was. Amazingly, though, when the experimenter had not seen the toy move, infants looked longer when she reached into the green box than into the yellow box – as if these 15-month-old infants expected her to reach for the toy where she falsely believed it to be.

Initially, these results were met with some skepticism, as they seem to contradict so much evidence that children do not understand false beliefs until many years later. However, in the intervening years, more and more studies of false belief understanding in infants and toddlers have accumulated. Baillargeon and colleagues have conducted a whole series of elegant studies, expanding on their original results. In these experiments, 1-year-olds (12–24 months) have demonstrated systematic expectations about actions based on false beliefs about contents as well as locations, based on indirect inferences as well as direct perception, and based on beliefs updated from other people's verbal reports as well as from observation (Onishi and Baillargeon, 2005; Scott and Baillargeon, 2009; Song et al., 2008; Surian et al., 2007).

Evidence that toddlers understand false beliefs is not restricted to studies using violation of expectation looking-time measures. Using anticipatory looking, Southgate et al. (2007) measured infants' predictions about where the experimenter would reach for her toy. If the experimenter had a false belief about her toy, 24-month-olds (but not 18-month-olds) anticipated that she would reach for her toy in the location where she saw it last, on the very first trial of the experiment.

Apparently, by their second year of life, children are already able to use inferred false beliefs to correctly predict others' actions. Why, then, is there a dramatic change in false belief task performance, when children are 4 years old? Baillargeon and colleagues suggest that 2- and 3-year-old children have a fully mature

understanding of representational mental states, but the demanding format of standard false belief tasks masks their abilities, rendering them incapable of expressing their knowledge (Baillargeon et al., 2010).

If so, the predictions for developmental cognitive neuroscience studies of ToM should be very different (Scott and Baillargeon, 2009). Brain regions specifically involved in ToM, like the TPJ and DMPFC, should not show any distinct qualitative developmental change around age 4 years, when children pass standard explicit false belief tasks. Instead, passing explicit false belief tasks should be associated with development in brain regions for executive function and language (see Botvinick et al., 2004; Caplan, 2007, for reviews). The developmental changes in ToM regions, by contrast, should occur much earlier, perhaps just after a child's first birthday.

Another interpretation of the infants' performance might be that while infants make systematic action predictions based on false beliefs, they do not actually have the same 'concept' of false beliefs that adults do. For example, infants may have a restricted, implicit understanding of beliefs, while 3–5-year-old children struggle to acquire a richer, more flexible, explicit concept of beliefs (Apperly and Butterfill, 2009). This view is plausible in part because a similar process occurs in other domains of cognition. The best-studied example is children's numerical concepts. Preverbal infants have representations with numerical content that allow them, for example, to track, differentiate, and even add and subtract small numbers (e.g., $1+1=2$; Feigenson et al., 2002; Wynn, 1992). Preverbal infants can also distinguish large numbers, when they differ by a large enough ratio (e.g., $8>4$, $16>8$; Xu and Spelke, 2000). However, these infants cannot form exact representations of large numbers that would let them, for example, distinguish between seven and eight objects (Carey, 2009; Xu and Spelke, 2000). Between the ages of 2 and 4 years, children then slowly and effortfully construct the concepts of the natural numbers, using as a necessary scaffold a culturally constructed list of names for numbers (Le Corre and Carey, 2007; Sarnecka and Gelman, 2004). The new concepts thus constructed are vastly more powerful than the infant's initial numerical conceptions.

It is thus tempting to believe that a similar process differentiates the infant's implicit conception of beliefs from the 4-year-old's explicit concept. Perhaps infants have an efficient, but limited, system for representing a person's belief about simple perceptual experiences (e.g., where she thinks the watermelon is). As they grow older, children gradually acquire an independent system that supports much richer representations that can be integrated with other processes to allow sophisticated explanations for others' behaviors, flexible revisions of the beliefs, and even moral judgments about others based on their beliefs. However, adherents of this view need to provide a characterization of the implicit versus explicit systems of ToM. What kind of competence does the implicit system support, and what are the key restrictions or limitations on the implicit conception? How does the explicit system overcome the limitations of the implicit one? In order to address this question, we need to go beyond the format of the tasks (i.e., looking-time or anticipatory looking versus explicit pointing or verbal responses) and focus on the nature of beliefs entertained by the two systems. Apperly and Butterfill (2009), for example, proposed that an implicit ToM might be limited to 'tracking attitudes to object's locations,' so infants might not be able to 'track beliefs involving both the features and the location of an object (e.g., 'the red ball is in the cupboard').' Since then, Baillargeon and colleagues have shown that infants can track this kind of belief too (e.g., the belief that 'the disassembled toy penguin is in the opaque box'; Scott and Baillargeon, 2009). However, this kind of proposal is precisely what is needed to give substance to the idea of two independent 'systems' for ToM.

Developmental cognitive neuroscience could make a key impact here. If implicit ToM and explicit ToM are truly distinct systems, they might have distinct neural mechanisms. The distinction might appear in the recruitment of different brain regions for implicit versus explicit ToM tasks. Alternatively, there might be two distinct temporal phases in the development of a single brain region (or group of regions): an early change corresponding to the initial development of an implicit ToM, and a later change when the same region is coopted for a qualitatively different explicit ToM. All of these possibilities remain to be tested.

20.2.2 Neuroimaging

Functional neuroimaging of children is a new research program, so very few fMRI studies have looked at the development of ToM brain regions. However, these few studies also pose challenges to the predictions of the Standard view. Specifically, the Standard view predicted that very young children rely on the pSTS representation of intentional actions, and that sometime around age 4 there is qualitative change in the neural representations of beliefs and desires. What the Standard model certainly did not predict was that the neural representations of both intentional action and ToM would show qualitative functional change in 5–10-year-old children. But that is exactly what the recent fMRI studies are finding.

The first demonstration of a neural response to intentional action in children was provided by Mosconi et al. (2005), who showed that the right pSTS in 7–10-year-old children is sensitive to the intentionality of action as in adults: while any gaze shift evokes activity in the right

pSTS, the activity is higher when the gaze shift is directed to unexpected locations than when the shift predictively follows a moving target. Interestingly, Carter and Pelphrey (2006) found a developmental change in the functional profile of this region much later than what the Standard view would predict. They measured brain activity in 7–10-year-old children while they viewed animated clips of biological (e.g., a person walking) and nonbiological motion (e.g., a grandfather clock moving). While biological motion induced higher activity than nonbiological motion in general, they found that this difference grows larger with age. These results suggest that although the right pSTS is already sensitive to intentional action before age 7, there are still ongoing changes in the specificity of response in this region well beyond this age (see Chapter 19). Therefore, the prediction of the Standard view that the right pSTS develops early in childhood before age 4 seems at least partly wrong.

Similarly, the few existing developmental fMRI studies on ToM (Kobayashi et al., 2007a, 2007b; Saxe et al., 2009) have all reported activations in similar brain regions in children that are shown to be recruited for ToM in adults: these areas include the bilateral TPJ, PC, and the MPFC. However, these studies also found developmental changes in the neural basis of ToM between school-age children and adults (Kobayashi et al., 2007a), or among school-age children (Saxe et al., 2009) in some of these regions.

In a recent study, Gweon et al. (2012) asked children between 5 and 12 years of age and adults to listen to stories inside the scanner and measured their brain responses to these stories. The stories described people's thoughts, beliefs, and feelings (mental condition), people's appearance and their social relationships (social condition), or purely physical events involving objects that do not involve people or their mental states (physical condition). The first question was whether children, just like adults, show higher responses to mental stories compared with physical stories. The answer was yes: the regions that were significantly more active to mental versus physical condition in children were strikingly similar to those found in adults (see Figure 20.1). However, there was an age-related change in how some these regions respond to the social stories. In the older half of the participants (8.5–12 years), the bilateral TPJ and PC did not respond to the social stories but only showed heightened activity to mental stories just like in adults. In contrast, the same brain regions in the younger half of the participants (5–8.5 years) responded to both mental and social stories, and did not discriminate stories with and without mental state content. That is, the selectivity of these regions to mental states increased with age. Saxe et al. (2009) found the same pattern (see Figure 20.4). In line with the studies on developmental change in the right pSTS, these results suggest that although the neural basis for ToM seems to be already in place before age 5, there are qualitative changes in the response selectivity of these brain areas that occur well past age 5. Furthermore, the selectivity in the RTPJ was correlated with children's performance outside the scanner on tasks designed to tap into later-developing aspects of ToM, such as making moral decisions based on mental states or understanding nonliteral utterances in context. The role of maturational factors and experience in the neural and behavioral development remains to be tested. However, these results provide initial evidence for a link between neural and cognitive development in ToM.

Rather than resolving the controversy between the Standard view and its more recent opponent, these results challenge both views and add another puzzle. The Standard view predicts that ToM brain regions

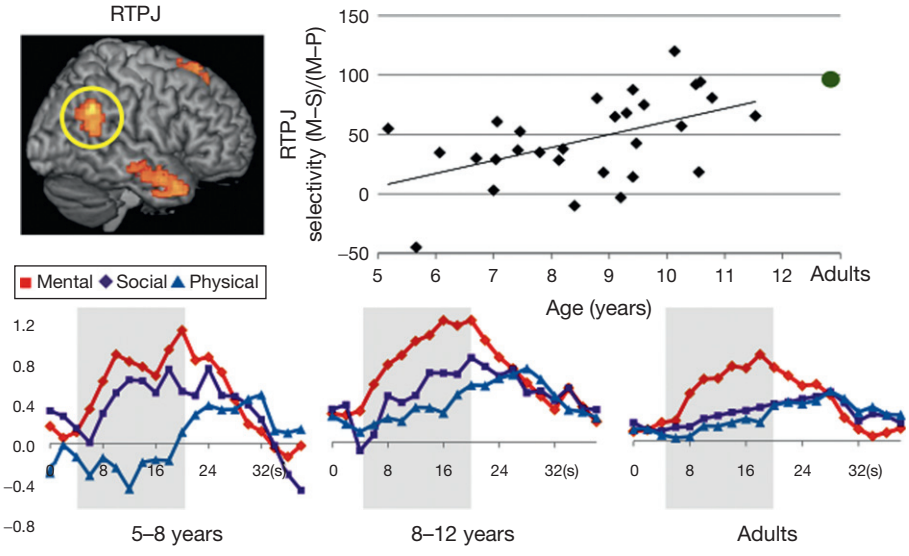

FIGURE 20.4 Developmental change in the RTPJ. The three graphs (bottom) show the time course of activation in the RTPJ in mental, social, and physical conditions in three different age groups (Gweon et al., 2012). The *X*-axis shows the time (seconds), and the *Y*-axis is the percent signal change in the activation relative to baseline. The stories were presented from 4 to 24 s, indicated by the shaded box. While the RTPJ in younger children responds to both mental and social conditions, the response to social conditions gets gradually lower as children grow older. That is, the RTPJ response becomes more 'selective' for mental stories. The scatterplot (upper right) shows that the selectivity of the RTPJ for mental, versus social, information becomes higher with age, in a cross-sectional sample of 32 children (combined data from Saxe et al., 2009; Gweon et al., 2012).

should show functional changes between the ages of 3 and 5 years. More recent data predict that significant changes in ToM brain regions should occur around the child's first birthday, when infants start to show signs of false belief understanding, with changes around age 4 years in regions supporting executive function. Therefore, the recent heated debates concerning the time course of development in the ToM mechanism have focused on these two age ranges (see Leslie, 2005; Perner and Ruffman, 2005; Scott and Baillargeon, 2009); neither predicted functional change in the neural basis of ToM well past 5 years of age.

Of course, these results do not provide direct evidence against either view, and these possibilities are not mutually exclusive. It remains possible that major changes occur in the ToM brain regions either around 12–15 months, or age 4 years, or both, supporting the acquisition of a concept of implicit and explicit false belief. However, if so, those hypothesized changes do not produce a brain region with a highly selective role in attributing mental states before 5 years of age: before age 8 years, children show normal brain regions involved in ToM (bilateral TPJ, PC, and regions in MPFC), but none of them are selectively recruited just for thinking about thoughts.

20.3 CONCLUSIONS

In sum, cognitive scientists have made foundational discoveries about the development of ToM in childhood, and about the neural basis of ToM in adulthood. Initially, these two independent methods of inquiry seemed to provide converging evidence of the structure of ToM: a domain-specific, human-unique mechanism that undergoes qualitative change in early childhood. Just recently, however, this unified view has begun to fracture, and much of what we thought we knew seems to be wrong.

The current evidence from behavioral studies of ToM suggests that infants have some understanding of false beliefs and how they affect actions, while 3-year-olds struggle to apply the same knowledge in similar tasks. Furthermore, while behavioral studies push the critical age for ToM development younger, neuroimaging studies suggest that the whole ToM system in the brain undergoes significant functional change much later, around age 8 years.

Of course, these puzzles may be partially due to the limitations in methods and the difficulty in using the same paradigm across different age groups. For example, standard false belief tasks have been mostly used for children past 3 years of age, whereas nonverbal false belief tasks using looking-time methods have been mostly used with infants and toddlers (but see Scott et al., 2012; Senju et al, 2009). This makes it difficult to directly test the differences in the nature of belief representations in infants and older children. Similarly, the available methods for measuring neural responses to these tasks vary by age. fMRI, which has been most extensively used for studies of ToM in adults and older children, is currently not useful for children younger than 4–5 years of age. In the future, it will be necessary to establish a clear relationship between fMRI results and those from other neuroimaging methods such as near infrared spectroscopy (NIRS), EEG, or magnetoencephalography (MEG), available for younger children.

The impact of a new understanding of the neural development of ToM could be tremendous. ToM deficit is a central issue in ASD. Knowing how brain regions for ToM emerge during typical development will provide the foundation for understanding how this development can go awry.

We are confident that the current uncertainty provides an important window of opportunity for new methods to make key theoretical contributions. We look forward to the emerging developmental cognitive neuroscience of ToM.

SEE ALSO

Cognitive Development: A Neuroscience perspective on empathy and its development; Developmental Neuroscience of Social Perception; Early Development of Speech and Language: Cognitive, Behavioral and Neural Systems; The Neural Architecture and Developmental Course of Face Processing; Theories in Developmental Cognitive Neuroscience. **Diseases:** Autisms.

Acknowledgments

The authors thank Jorie Koster-Hale for comments on the manuscript and Shawn Green for help with the figures. R Saxe was supported by the John Merck Scholars Program, the Simons Foundation, the Packard Foundation, and the Ellison Medical Foundation.

References

Apperly, I., Butterfill, S., 2009. Do humans have two systems to track beliefs and belief-like states. Psychological Review 116, 953–970.

Avis, J., Harris, P., 1991. Belief-desire reasoning among Baka children: Evidence for a universal conception of mind. Child Development 62, 460–467.

Baillargeon, R., Scott, R., He, Z., 2010. False-belief understanding in infants. Trends in Cognitive Sciences 14, 110–118.

Baron-Cohen, S., 1997. Mindblindness: An Essay on Autism and Theory of Mind. The MIT Press, Cambridge, MA.

Baron-Cohen, S., Leslie, A.M., Frith, U., 1985. Does the autistic child have a "theory of mind"? Cognition 21, 37–46.

Bedny, M., Pascual-Leone, A., Saxe, R., 2009. Growing up blind does not change the neural bases of theory of mind. Proceedings of the National Academy of Sciences 106, 11312–11317.

Botvinick, M., Cohen, J., Carter, C., 2004. Conflict monitoring and anterior cingulate cortex: An update. Trends in Cognitive Sciences 8, 539–546.

Brass, M., Schmitt, R.M., Spengler, S., Gergely, G., 2007. Investigating action understanding: Inferential processes versus action simulation. Current Biology 17, 2117–2121.

Brunet, E., Sarfati, Y., Hardy-Baylè, M., Decety, J., 2000. A PET investigation of the attribution of intentions with a nonverbal task. NeuroImage 11, 157–166.

Callaghan, T., Rochat, P., Lillard, A., et al., 2005. Synchrony in the onset of mental-state reasoning. Psychological Science 16, 378–384.

Caplan, D., 2007. Functional neuroimaging studies of syntactic processing in sentence comprehension: A critical selective review. Language and Linguistics Compass 1, 32–47.

Carey, S., 2009. The Origin of Concepts. Oxford University Press Inc., New York.

Carter, E., Pelphrey, K., 2006. School-aged children exhibit domain-specific responses to biological motion. Social Neuroscience 1, 396–411.

Castelli, F., Happe, F., Frith, U., Frith, C., 2000. Movement and mind: A functional imaging study of perception and interpretation of complex intentional movement patterns. NeuroImage 12, 314–325.

Charman, T., Baron-Cohen, S., 1992. Understanding drawings and beliefs: A further test of the metarepresentation theory of autism: A research note. Journal of Child Psychology and Psychiatry 33, 1105–1112.

Downing, P.E., Jiang, Y., Shuman, M., Kanwisher, N., 2001. A cortical area selective for visual processing of the human body. Science 293, 2470–2473.

Feigenson, L., Carey, S., Hauser, M., 2002. The representations underlying infants' choice of more: Object files versus analog magnitudes. Psychological Science 13, 150–156.

Fletcher, P.C., Happe, F., Frith, U., et al., 1995. Other minds in the brain: A functional imaging study of "theory of mind" in story comprehension. Cognition 57, 109–128.

Gallagher, H.L., Happe, F., Brunswick, N., Fletcher, P.C., Frith, U., Frith, C.D., 2000. Reading the mind in cartoons and stories: An fMRI study of 'theory of mind' in verbal and nonverbal tasks. Neuropsychologia 38, 11–21.

Gergely, G., Csibra, G., 2003. Teleological reasoning in infancy: The naive theory of rational action. Trends in Cognitive Sciences 7, 287–292.

German, T.P., Niehaus, J.L., Roarty, M.P., Giesbrecht, B., Miller, M.B., 2004. Neural correlates of detecting pretense: Automatic engagement of the intentional stance under covert conditions. Journal of Cognitive Neuroscience 16, 1805–1817.

Gobbini, M.I., Koralek, A.C., Bryan, R.E., Montgomery, K.J., Haxby, J.V., 2007. Two takes on the social brain: A comparison of theory of mind tasks. Journal of Cognitive Neuroscience 19:11, 1803–1814.

Goel, V., Grafman, J., Sadato, N., Hallett, M., 1995. Modeling other minds. Neuroreport 6, 1741–1746.

Goodman, N.D., Baker, C.L., Bonawitz, E.B., et al., 2006. Intuitive theories of mind: A rational approach to false belief. In: Proceedings of the Twenty-Eighth Annual Conference of the Cognitive Science Society, Vancouver, Canada.

Gopnik, A., 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.

Grossman, E., Donnelly, M., Price, R., et al., 2000. Brain areas involved in perception of biological motion. Journal of Cognitive Neuroscience 12, 711–720.

Gweon, H., Dodell-Feder, D., Bedny, M., and Saxe, R., 2012. Theory of mind performance in children correlates with functional specialization of a brain region for thinking about thoughts. *Child Development* doi:10.1111/j.1467-8624.2012.01829.x.

Happe, F., Ehlers, S., Fletcher, P., et al., 1996. 'Theory of mind' in the brain. Evidence from a PET scan study of Asperger syndrome. Neuroreport 8, 197.

Howard, R., Brammer, M., Wright, I., Woodruff, P., Bullmore, E., Zeki, S., 1996. A direct demonstration of functional specialization within motion-related visual and auditory cortex of the human brain. Current Biology 6, 1015–1019.

Kobayashi, C., Glover, G.H., Temple, E., 2007a. Children's and adults' neural bases of verbal and nonverbal 'theory of mind'. Neuropsychologia 45, 1522–1532.

Kobayashi, C., Glover, G.H., Temple, E., 2007b. Cultural and linguistic effects on neural bases of 'theory of mind' in American and Japanese children. Brain Research 1164, 95–107.

Le Corre, M., Carey, S., 2007. One, two, three, four, nothing more: An investigation of the conceptual sources of the verbal counting principles. Cognition 105, 395–438.

Leekam, S.R., Perner, J., 1991. Does the autistic child have a metarepresentational deficit? Cognition 40, 203–218.

Leslie, A.M., 2005. Developmental parallels in understanding minds and bodies. Trends in Cognitive Sciences 9, 459–462.

Leslie, A.M., Thaiss, L., 1992. Domain specificity in conceptual development: Neuropsychological evidence from autism. Cognition 43, 225–251.

Meltzoff, A.N., Broks, R., 2008. Self-experience as a mechanism for learning about others: A training study in social cognition. Developmental Psychology 44 (5), 1257–1265.

Mosconi, M., Mack, P., McCarthy, G., Pelphrey, K., 2005. Taking an "intentional stance" on eye-gaze shifts: A functional neuroimaging study of social perception in children. NeuroImage 27, 247.

Moses, L., Flavell, J., 1990. Inferring false beliefs from actions and reactions. Child Development 61, 929–945.

Onishi, K.H., Baillargeon, R., 2005. Do 15-month-old infants understand false beliefs? Science 308, 255–258.

Parkin, L., 1994. Children's Understanding of Misrepresentation. University of Sussex, UK (unpublished manuscript).

Pascual-Marqui, R.D., Esslen, M., Kochi, K., Lehmann, D., 2002. Functional imaging with low-resolution brain electromagnetic tomography (LORETA): A review. Methods and Findings in Experimental and Clinical Pharmacology 24, 91–95.

Pelphrey, K., Singerman, J., Allison, T., McCarthy, G., 2003. Brain activation evoked by perception of gaze shifts: The influence of context. Neuropsychologia 41, 156–170.

Pelphrey, K.A., Morris, J.P., McCarthy, G., 2004. Grasping the intentions of others: The perceived intentionality of an action influences activity in the superior temporal sulcus during social perception. Journal of Cognitive Neuroscience 16, 1706–1716.

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

Perner, J., Ruffman, T., 2005. Infants' insight into the mind: How deep? Science 308, 214.

Perner, J., Aichorn, M., Kronblicher, M., Staffen, W., Ladurner, G., 2006. Thinking of mental and other representations: The roles of right and left temporo-parietal junction. Social Neuroscience 1, 245–258.

Peterson, C., Wellman, H., Liu, D., 2005. Steps in theory-of-mind development for children with deafness or autism. Child Development 76, 502–517.

Phillips, A.T., Wellman, H.M., 2005. Infants' understanding of object-directed action. Cognition 98, 137–155.

Sabbagh, M., Bowman, L., Evraire, L., Ito, J., 2009. Neurodevelopmental correlates of theory of mind in preschool children. Child Development 80, 1147–1162.

Sarnecka, B., Gelman, S., 2004. Six does not just mean a lot: Preschoolers see number words as specific. Cognition 92, 329–352.

Saxe, R., Kanwisher, N., 2003. People thinking about thinking people. The role of the temporo-parietal junction in "theory of mind" NeuroImage 19, 1835–1842.

Saxe, R., Powell, L.J., 2006. It's the thought that counts: Specific brain regions for one component of theory of mind. Psychological Science 17, 692–699.

Saxe, R., Wexler, A., 2005. Making sense of another mind: The role of the right temporo-parietal junction. Neuropsychologia 43, 1391–1399.

Saxe, R., Carey, S., Kanwisher, N., 2004a. Understanding other minds: Linking developmental psychology and functional neuroimaging. Annual Review of Psychology 55, 87–124.

Saxe, R., Xiao, D.K., Kovacs, G., Perrett, D.I., Kanwisher, N., 2004b. A region of right posterior superior temporal sulcus responds to observed intentional actions. Neuropsychologia 42, 1435–1446.

Saxe, R., Schulz, L.E., Jiang, Y.V., 2006. Reading minds versus following rules: Dissociating theory of mind and executive control in the brain. Social Neuroscience 1, 284–298.

Saxe, R., Whitfield-Gabrieli, S., Scholz, J., Pelphrey, K.A., 2009. Brain regions for perceiving and reasoning about other people in school-aged children. Child Development 80, 1197–1209.

Scott, R., Baillargeon, R., 2009. Which penguin is this? Attributing false beliefs about object identity at 18 months. Child Development 80, 1172–1196.

Scott, R., He, Z., Baillargeon, R., Cummins, D., 2012. False-belief understanding in 2.5-year-olds: Evidence from two novel verbal spontaneous-response tasks. Developmental Science 15, 181–193.

Senju, A., Southgate, V., White, S., Frith, U., 2009. Mindblind eyes: An absence of spontaneous theory of mind in Asperger syndrome. Science 325 (5942), 883–885.

Song, H., Onishi, K., Baillargeon, R., Fisher, C., 2008. Can an actorís false belief be corrected by an appropriate communication? Psychological reasoning in 18.5-month-old infants. Cognition 109, 295–315.

Southgate, V., Senju, A., Csibra, G., 2007. Action anticipation through attribution of false belief by 2-year-olds. Psychological Science 18, 587–592.

Surian, L., Caldi, S., Sperber, D., 2007. Attribution of beliefs by 13-month-old infants. Psychological Science 18, 580–586.

Thatcher, R., 1992. Cyclic cortical reorganization during early childhood. Brain and Cognition 20, 24–50.

Tootell, R., Reppas, J., Kwong, K., et al., 1995. Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. Journal of Neuroscience 15, 3215–3230.

Vogeley, K., Bussfeld, P., Newen, A., et al., 2001. Mind reading: Neural mechanisms of theory of mind and self-perspective. NeuroImage 14, 170–181.

Wagner, M., Fuchs, M., Kastner, J., 2004. Evaluation of sLORETA in the presence of noise and multiple sources. Brain Topography 16, 277–280.

Wellman, H.M., Cross, D., Watson, J., 2001. Meta-analysis of theory-of-mind development: The truth about false belief. Child Development 72, 655–684.

Wimmer, H., Perner, J., 1983. Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. Cognition 13, 103–128.

Woodward, A.L., 1998. Infants selectively encode the goal object of an actor's reach. Cognition 69, 1–34.

Wynn, K., 1992. Addition and subtraction by human infants. Nature 358, 749–750.

Xu, F., Spelke, E., 2000. Large number discrimination in 6-month-old infants. Cognition 74, B1–B11.

Zaitchik, D., 1990. When representations conflict with reality: The preschooler's problem with false beliefs and "false" photographs. Cognition 35, 41–68.