

UNDERSTANDING OTHER MINDS: Linking Developmental Psychology and Functional Neuroimaging

R. Saxe, S. Carey, and N. Kanwisher

*Department of Brain and Cognitive Sciences, MIT, Cambridge, Massachusetts 02139;
email: saxe@mit.edu, scarey@wjh.harvard.edu, ngk@mit.edu*

Key Words theory of mind, mentalizing, social cognition, cognitive neuroscience

■ **Abstract** Evidence from developmental psychology suggests that understanding other minds constitutes a special domain of cognition with at least two components: an early-developing system for reasoning about goals, perceptions, and emotions, and a later-developing system for representing the contents of beliefs. Neuroimaging reinforces and elaborates upon this view by providing evidence that (a) domain-specific brain regions exist for representing belief contents, (b) these regions are apparently distinct from other regions engaged in reasoning about goals and actions (suggesting that the two developmental stages reflect the emergence of two distinct systems, rather than the elaboration of a single system), and (c) these regions are distinct from brain regions engaged in inhibitory control and in syntactic processing. The clear neural distinction between these processes is evidence that belief attribution is not dependent on either inhibitory control or syntax, but is subserved by a specialized neural system for theory of mind.

CONTENTS

INTRODUCTION	88
FUNCTIONAL NEUROIMAGING: STANDARDS OF EVIDENCE AND INFERENCE	89
DEVELOPMENTAL PSYCHOLOGY: BELIEF ATTRIBUTION	93
Inhibitory Control and Belief Attribution	94
Language and Belief Attribution	95
Beyond False Belief	96
NEUROIMAGING: BELIEF ATTRIBUTION	97
Neuroimaging: Belief Attribution and Inhibitory Control	100
Neuroimaging: Belief Attribution and Language	101
Belief Attribution: Conclusions	103
DEVELOPMENTAL PSYCHOLOGY: DESIRES, PERCEPTIONS, EMOTIONS	103

Desires/Goals	104
Neuroimaging: Desires, Perceptions, and Emotions	108
CONCLUSIONS	115

INTRODUCTION

Unlike behaviorists, normal adults attribute to one another (and to themselves) unobservable internal mental states, such as goals, thoughts, and feelings, and use these to explain and predict behavior. This human capacity for reasoning about the mental causes of action is called a theory of mind. In the past 25 years, theory of mind has become a major topic of research, initially in developmental psychology and subsequently in other fields including social psychology, philosophy, and ethology.

Most recently, a new method for theory of mind research has joined the pack: functional brain imaging [especially functional magnetic resonance imaging (fMRI)]. The blood oxygenation level dependent (BOLD) signal measured by fMRI gives scientists unprecedented access to the hemodynamic changes (and indirectly to the neural activity) in the brain that are associated with psychological processes. fMRI may have by now exceeded all other techniques in psychology in terms of expense, growth rate, and public visibility. But many have wondered how much this new technology has actually contributed to the study of human cognition, and when—if ever—a finding from functional neuroimaging has constrained a cognitive theory.

In this review, we take the human theory of mind as a case study of real theoretical exchange between studies (and scientists) using functional neuroimaging and those using the more well-established techniques of developmental psychology. As such, this review will necessarily be selective. For more complete coverage of related research, we refer the reader to the many excellent recent reviews in the fields of cognitive neuroscience (e.g., Adolphs 2001, 2002, 2003; Allison et al. 2000; Blakemore & Decety 2001; Decety & Grezes 1999; Frith 2001; Frith & Frith 2000, 2003; Gallagher & Frith 2003; Gallese 2003; Greene & Haidt 2002; Grezes & Decety 2001; Puce & Perrett 2003; Siegal & Varley 2002) and developmental psychology (e.g., Baldwin & Baird 2001; Bartsch 2002; Csibra 2003; Flavell 1999; Johnson 2000, 2003; Meltzoff & Decety 2003; Wellman & Lagattuta 2000; Wellman et al. 2001).

This review is divided into two main sections, following the substantial evidence for at least two distinct stages in the development of theory of mind (see Figure 1*a*). The first half of the review deals with belief attribution. At approximately age 3 or 4 children begin to attribute representational epistemic mental states—thoughts, beliefs, and knowledge—to themselves and others. The second half of the review considers the earlier-developing mentalistic reasoning that occurs before age 3. Toddlers do reason about the mind and human behavior, but they do so with a more limited repertoire of mental state concepts, including desires,

perceptions, and emotions. Within each section of the review, we first summarize theoretical questions and findings emerging from developmental psychology, and then consider potential and actual contributions made by functional neuroimaging to answering these questions.

Functional neuroimaging is particularly well suited to resolve questions of whether two tasks or processes engage common or distinct mechanisms. For instance, it has been suggested that the development of a concept of belief depends critically on the ability to represent sentence-complement syntax (e.g., de Villiers 2000). However, this dependence could arise either because children cannot learn about beliefs until they can understand the language adults use to talk about the mind, or it could arise because belief attribution is truly dependent on the cognitive and neural mechanisms for parsing sentence complement syntax (even in adulthood). Functional neuroimaging allows us to ask whether these two tasks recruit the same or different regions of the brain.¹ If different brain regions are involved in belief attribution and syntax, then it is less likely that a single functional mechanism is responsible for both. While this use of neuroimaging is potentially powerful in answering fundamental questions about cognition, it is subject to several pitfalls and ambiguities that are discussed briefly in the next section.

FUNCTIONAL NEUROIMAGING: STANDARDS OF EVIDENCE AND INFERENCE

If we are to accept a finding that two different tasks activate the same brain region as evidence that common psychological mechanisms are engaged in the two tasks, then we must consider two questions. First, what counts as the same brain region, and what kind of data can support a claim of common or distinct activations? Second, what is the relationship between brain regions and psychological mechanisms?

The location of an activation in the brain is often specified by general region (e.g., the occipital pole, or the temporo-parietal junction), or by the gyrus or sulcus where the activation is found (e.g., the fusiform gyrus, or the intraparietal sulcus). These descriptors can be useful, but are not very precise, as each one spans ten

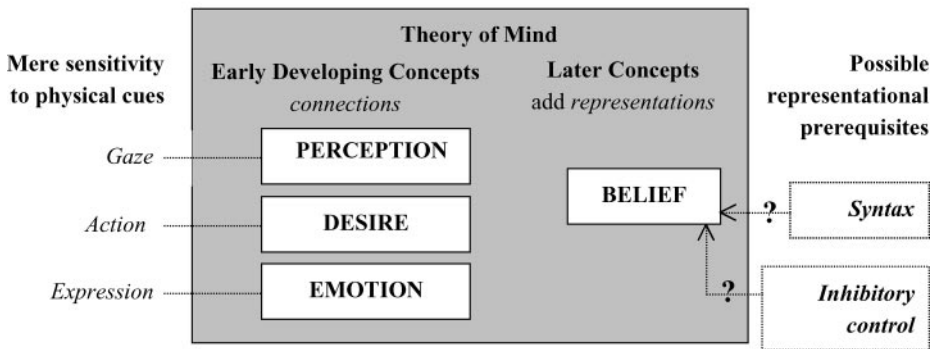
¹All of the functional imaging discussed in this paper used human adult subjects. In some cases, the adult results alone are sufficient to constrain theories, as this review illustrates. However, other questions and hypotheses could be tested only in the brains of infants and children themselves. For instance, are the brain regions involved in response conflict and in belief attribution independent in young children just beginning to attribute beliefs, as they are in adults? fMRI in children and infants is possible (e.g., Born et al. 1998, Burgund et al. 2002, Casey et al. 2001, Dehaene-Lambertz et al. 2002), but is not widespread. In particular, there are no published fMRI studies yet of children between 10 and 48 months old, the critical age range for theory of mind development, and no fMRI studies of theory of mind in any pediatric population.

or more square centimeters of cortex. Such large regions are likely to encompass many functionally distinct areas, as seen, for example, in extrastriate cortex where areas such as the visual motion area MT (Tootell et al. 1995) or the fusiform face area (Kanwisher et al. 1997, McCarthy et al. 1997) are typically one or two square centimeters in size. Thus, even if two tasks produce activations within the same general region of the brain, their activations may not overlap at all. (Imagine concluding that the brain does not contain distinct motor representations of hands and feet because both hand and foot movements are coordinated by primary motor cortex.) A further problem is that because individual brains differ from each other physically, there is no theory-neutral way to precisely specify what counts as the “same place” in two different brains.

These problems can be avoided in analyses of individual subjects. The strongest evidence for engagement of the same brain region by two different tasks arises

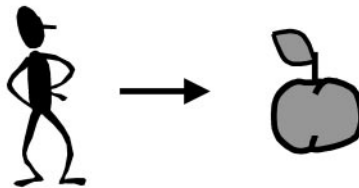
(a)

Attributing mental states to other minds



(b)

Connection:



Representation:



when the very same voxels in the same subject's brain (preferably from the same scanning session) are significantly activated by each of two different tasks. While some neuroimaging studies meet this high standard, many do not.

A related approach is to first functionally define a region of interest (ROI) individually within each subject based on a particular task comparison (or localizer scan), then pool over the voxels in that ROI to ask whether a second task activates the same ROI. This approach avoids the problem of having to register different brains, while making possible statistical analyses over multiple subjects' ROIs. However, in choosing a localizer to define an ROI the researcher is making an ontological assumption that this localizer contrast picks out a meaningful functional unit in the brain (i.e., a natural kind). Like other ontological assumptions

Figure 1 Schematic representation of the two principle stages in the development of theory of mind, and the central theoretical questions discussed in this review. (*a, left*) Toddlers reason about other minds with a limited repertoire of mental state concepts, including desire/goal, perception, and emotion. For both developmental psychology and neuroimaging, it is critical to distinguish attribution of desires, perceptions, and emotions (the "rich" or "mentalistic" interpretation) from mere behavioral sensitivity to the associated physical cues (including human body motions, gaze direction and emotional expressions—the "lean" interpretation). Two important characteristics of mentalistic attribution are reference (mental states are about objects or situations) and coherence (different mental state attributions interact causally and systematically). It is an open question, addressed in this review, whether these three mental state concepts are attributed using one common mechanism or multiple distinct mechanisms. (*a, right*) Starting at approximately age 3 or 4 children include a concept of belief in their reasoning about other minds. A central debate in developmental psychology concerns whether this later development reflects true conceptual change in the child's theory of mind, or simply the maturation of other capacities that are necessary for reasoning about beliefs. Two prominent candidate capacities are syntax and inhibitory control. Evidence reviewed here suggests that representing others' beliefs does not recruit the same brain mechanisms as either syntax or inhibitory control. (*b*) What then is the nature of the conceptual change between a toddler's theory of mind, and the later theory that incorporates attributions of belief? One possibility is that the toddler lacks the notion of a representational mental state, and instead conceives of mental relations between people and the world as direct connections. We can think of a connection as something like gravity or resonance to affordances in the environment, leaving no possibility for error or misperception. Thus goals, perceptions, and emotions may at first be understood as referential (about an object or situation) but not as representational (requiring an independent representation of the object or situation in the mind of the actor). The terms "connection" versus "representation" were used to characterize this developmental change by Flavell (1988), but a similar distinction is included in the theories of Perner (1993) and Wellman (e.g., Bartsch & Wellman 1995). The figure is adapted with permission from Bartsch & Wellman (1995).

in science, the utility of a particular functionally defined ROI is determined by the consistency of the data that emerge from it and the richness of the theoretical progress those data support.

Most common in the neuroimaging literature are group analyses, in which brain images from a dozen or more individuals are aligned (as best as possible) into a common space, and statistical analyses are then conducted across subjects. This method enables one to test whether an activation pattern is consistent across subjects, and it can sometimes provide greater statistical power than individual-subject analyses. However, it comes at the cost of blurring of activation maps due to the necessarily imperfect registration across physically different brains. Individuals vary not only in their physical anatomy but also in their functional anatomy, producing yet more blurring in group-averaged data. Thus, activations that may be completely nonoverlapping within each individual could be highly overlapping when the same data are averaged across subjects. This problem is exacerbated when comparing activations across subject groups or across studies.

Thus, although group-analyzed data can be informative, any claim that two tasks activate the same place in the brain are most convincing when they are based on analyses within individual subjects or within individually defined ROIs. The least convincing evidence for common mechanisms comes when each of two tasks produces an activation somewhere within the same large anatomical region (e.g., the temporo-parietal junction or superior temporal sulcus).

The opposite inference, that distinct cortical regions are engaged by two different tasks, is subject to a different and surprisingly common error. Researchers often argue that two tasks engage distinct mechanisms because task A activates region X significantly (compared to a control condition), whereas task B does not. Such arguments are not valid on their own: a difference in significances is not a significant difference. To argue for differential activation of region X by task A and task B, it is necessary to directly compare the activation for the two tasks.

Even with the proper statistical evidence, difficult theoretical issues remain. Psychological theories concern psychological processes whereas neuroimaging data can only test the activation of cortical voxels. What is the relationship between the two? A typical voxel in a neuroimaging study contains hundreds of thousands of neurons, so the common activation of a voxel by two different tasks could arise even if completely distinct neural populations within that voxel are engaged by the two tasks. Thus caution is required when we infer common mechanisms from common activations. That said, if a whole cluster of adjacent voxels shows activation for each of two different tasks but not for many others, it is a reasonable guess that even if distinct neural populations are involved, they are likely to be functionally related.

Caution is also required when making the opposite kind of inference. If each of two tasks activates a distinct and nonoverlapping cortical region, it is clear that different physical brain hardware is engaged by the two tasks, but this need not imply that qualitatively different psychological processes are involved in the two tasks. Consider a stimulus falling in the upper visual field versus the lower visual

field: distinct and nonoverlapping regions within primary visual cortex would be activated, but the kind of processing that goes on in each is presumably very similar. Deciding whether dissociations at the cortical level correspond to qualitative differences in processing at the psychological level is a difficult judgment call that requires consideration of the wider theoretical context (e.g., does it make sense computationally that different processes may be engaged by these two tasks?) and the available empirical evidence (e.g., what other tasks activate these two regions?).

Armed with these cautionary notes about the use of neuroimaging data to resolve psychological questions, we turn now to a review of the literature on the development of theory of mind.

DEVELOPMENTAL PSYCHOLOGY: BELIEF ATTRIBUTION

An understanding of other people's thoughts and beliefs plays a central role in adult reasoning about the causes of other people's behavior (e.g., "she spilled the coffee because she thought the cup was empty," "he's going home because he thinks he left his keys there") (Malle 2001), but not, of course, in reasoning about the causes of mechanical events like clockwork or the tides. Given the centrality of beliefs in adult explanations of actions, we might expect that as soon as children begin to explain their own and others' actions, they would use this powerful notion of belief to do so. On the contrary, it is now well established that children do not begin to use beliefs to explain actions until relatively late in development, at age 3 or 4.

The most common test of children's ability to explain an action with reference to the actor's belief is the "false belief" task (Wimmer & Perner 1983; for reviews of this literature see Flavell 1999, Wellman & Lagattuta 2000, Wellman et al. 2001). In the standard version of this task (the "object transfer" problem), the child is told a story in which a character's belief about the location of a target object becomes false when the object is moved without the character's knowledge. In Wimmer & Perner's original version, for instance, Maxi's mother moves the chocolate from the green to the blue cupboard while Maxi is outside playing. The children are then variously asked to report the content of the character's belief ("Where does Maxi *think* the chocolate is?"), to predict the character's action ("Where will Maxi *look* for the chocolate?"), or sometimes to explain the completed action ("Why did Maxi look for the chocolate in the *green* cupboard?"). The critical feature of a false belief task is that the correct answers to all three of these questions—even the ones that do not specifically query a belief content—require the child to pay attention to Maxi's belief, and not to the actual location of the chocolate (Dennett 1978, Premack & Woodruff 1978). Dozens of versions of the false belief problem have now been used, and while the precise age of success varies between children and between task versions (Wellman et al. 2001), in general children younger than 3 or 4 do not correctly solve false belief problems, but older children do.

The contentious issue is not when success on false belief problems emerges but what such success reflects. One possibility is that children 3 or 4 years old undergo real change in the concepts they use to reason about other minds, acquiring

a previously absent representational concept of belief (Flavell 1999, Perner 1993, Wellman et al. 2001), and producing a consequent improvement on false belief tasks.

A second possibility is that the concept of belief is already intact in young children but is masked in the false belief paradigm by immaturity in other capacities that are necessary for good performance on the task. Two such candidate capacities are inhibitory control and some aspects of syntactic knowledge (especially complement syntax), both of which are correlated with false belief task performance (Astington & Jenkins 1995, 1999; Carlson & Moses 2001; de Villiers & Pyers 2002; Watson et al. 2001). In the next two sections we will review the evidence for an association between the development of reasoning about beliefs and inhibitory control or syntax. We will focus in particular on evidence for (and against) false belief task performance as a measure of a newly acquired representational concept of belief, as opposed to simple unmasking by maturation of these other capacities.

Inhibitory Control and Belief Attribution

To answer a false belief question correctly, a child must be able to juggle two competing representations of reality (the actual state of affairs and the reality represented in the protagonist's head) and to inhibit an incorrect but compelling answer (the true location of the object). In variants of the false belief task, the demand for inhibitory control predicts children's performance. For instance, the current location of the target object may be made less salient and thus easier to inhibit: instead of being moved to the green cupboard, Maxi's chocolate is eaten, or the actual location of the chocolate is unknown to the child (described in Wellman et al. 2001; see also Zaitchik 1991). Children of all ages perform better on these versions of the task (Wellman et al. 2001). Conversely, when the inhibitory demands are increased by changing the protagonist's motivation to a negative desire (i.e., the protagonist's desire is not to find, but to *avoid* the target object), 4- and even 6-year-olds consistently fail to answer correctly (Leslie 2000, Leslie & Polizzi 1998). Finally, the inhibitory demands of false belief tasks are not restricted to reasoning about beliefs. Four-year-olds have more difficulty with logically equivalent problems about nonmental false representations (e.g., false photographs or maps; Leslie & Thaiss 1992, Zaitchik 1990), and with juggling two different verbal labels of a single object (Apperly & Robinson 2002).

Nevertheless the interpretation that false belief performance is limited by immature inhibitory control remains controversial (e.g., Perner et al. 1999). Wellman et al. (2001) note that the improvement in three-year-olds' performance on false belief tasks with salience manipulations reflects only a change from below-chance to chance performance, and therefore does not implicate an operational concept of belief (but see Moses 2001). They argue that the correlation between inhibitory control and false belief performance need not reflect masking of a preexisting competence. Instead, inhibitory control could facilitate knowledge acquisition and conceptual change in the domain of other minds, since a child who can disengage

from the prepotent representation of reality may be more able to focus on and learn about mental representations (Moses 2001, Wellman et al. 2001).

In all, the role of executive function or inhibitory control in reasoning about beliefs remains open to investigation (see also Perner & Lang 2000). Is inhibitory control recruited during successful false belief task performance? Is it recruited even more generally during all reasoning about beliefs? Is the locus of such inhibition predominantly peripheral (resolving response conflict) or cognitive (inhibiting prepotent representations)? Alternatively, if executive function contributes only to children's early learning about the mind, then we would not expect the same brain regions to be recruited when adults engage in belief attribution and inhibitory control. Below, we examine whether neuroimaging of healthy human adults can help resolve some of these questions.

Language and Belief Attribution

A striking demonstration of the role of language in false belief task performance comes from studies of deaf children. Deaf children of hearing parents (i.e., non-native signers) are impaired on sign language versions of false belief tasks (de Villiers & de Villiers 2000a, Peterson & Siegal 1995). This deficit persists even on nonverbal (i.e., pictorial) versions of the false belief task (de Villiers & de Villiers 2000a, Woolfe et al. 2002). Deaf children of native signers show no impairment. Conversely, the quantity and quality of family talk about mental causes to which a normally developing toddler is exposed is correlated with performance on false belief tasks two years later (e.g., Cutting & Dunn 1999).

The causal relationship between developing competences in language and "theory of mind" has been controversial, though. Some models propose that linguistic competence is a necessary precursor of theory of mind (e.g., de Villiers 2000) while others suggest that theory of mind is a necessary precursor of language development (e.g., Baron-Cohen et al. 1985, Bloom 2000, Happe 1992). One possibility is that early-developing components of theory of mind (discussed in detail below) are necessary for some aspects of language acquisition (e.g., establishing the referents of newly heard words), whereas other aspects of language acquisition such as the syntax of complementation and the semantics of opacity are in turn necessary for the late-developing concept of belief (Malle 2003). The relationship between belief attribution and language is addressed here.

The specific attribute of language commonly implicated in representing another person's beliefs is the syntax and semantics of sentential complements. de Villiers (e.g., 2000; de Villiers & de Villiers 2000a,b) has proposed that "language is the only representational system that could" support the concept of (false) beliefs, because language is "propositional, and can therefore capture falsity and embeddedness of propositions." Mental state verbs share with verbs of communication a particular syntactic structure of referentially opaque embedded complements (the truth value of the sentence is independent of the truth value of the complement, as in "John thinks that *it is raining*"). Children's production and comprehension of

this syntactic structure precedes and strongly predicts performance on both standard and nonverbal versions of the false belief task (de Villiers 2000; de Villiers & de Villiers 2000a,b; but see Bartsch & Wellman 1995, Ruffman et al. 2003).

However, linguistic skills may correlate with reasoning about beliefs simply because language enables a child to learn about (or to learn to talk about) the mind. Conversational experience (including comprehension of embedded sentence complements) contributes to children's developing knowledge about the mind, because first-person verbal report is our dominant source of information about subjective states occurring inside someone else's head (Harris 1989, Nelson 1996). Thus, sophisticated syntax may support the development of a concept of belief, but may not be recruited during reasoning with that concept.

What then is the role of language in adult reasoning about beliefs? If constructing the syntax of embedded propositional structures is necessary for all reasoning about (false) beliefs, then the same neural structures should be recruited by tasks that tap these two processes. This prediction is addressed below.

Beyond False Belief

In all, the extensive literature on false belief task performance in normally developing children has produced a consistent pattern of results and many competing interpretations. Wellman et al. (2001) thus conclude their meta-analysis of this literature with a plea that researchers move on to seek new and converging evidence for competing theories of the developing concept of belief. In fact, Bloom & German (2000) specify two reasons that researchers should abandon the false belief task as the benchmark of mature belief attribution. First, there is more to success on the false belief task than a concept of belief, as illustrated by the preceding review. Second, they point out, there is more to a concept of belief than passing the false belief task. In fact, theory of mind reasoning would not work if we did not attribute to others mostly true beliefs [and mostly rational actions (Dennett 1996)]. For all of these reasons, the false belief task is at best a limited tool for measuring the developing concept of belief.

So, is there evidence for the distinct emergence of a concept of belief, beyond the false belief task? The best such evidence is Bartsch & Wellman's (1995) investigation of children's spontaneous talk about beliefs. Bartsch & Wellman distinguish mere conversational turns of phrase ("know what?") from genuine psychological references, often identified by the children's use of contrastives: sentences using "think" or "know" that contrast expectations and outcomes, fiction and reality, or differences between individuals. The first genuine references to thoughts and beliefs appear around the third birthday, significantly later than genuine psychological reference to desires and emotions, but about half a year before children spontaneously explain actions in terms of beliefs (Bartsch & Wellman 1995), or pass false belief tasks (e.g., Wellman et al. 2001).

Evidence from both experimental tasks and spontaneous speech thus converge on a change in children's reasoning about beliefs that occurs in the third or fourth

year. Is this development reflected in a specialized neural substrate for reasoning about beliefs? If so, what is the relationship between this neural substrate and brain regions subserving inhibitory control? Language? These questions could be addressed using functional neuroimaging.

NEUROIMAGING: BELIEF ATTRIBUTION

The first question for neuroimaging is, Can we find regions of the adult human brain that show activity specifically when subjects are required to attribute beliefs to another person? Of the neuroimaging studies that attempt to address this question directly, four have followed developmental psychology in using false belief problems (verbal and nonverbal) as the definitive belief attribution task (Fletcher et al. 1995, Gallagher et al. 2000, Saxe & Kanwisher 2003, Voegeley et al. 2001). One study gave subjects simple descriptions of events involving people, and instructed subjects to “try to understand their motivations, feelings and actions” (Ferstl & von Cramon 2002). Goel et al. (1995) used an original task: Subjects were asked to judge whether Christopher Columbus could have identified the function of a pictured object. Across these studies, when subjects reason about false beliefs or Columbus’s knowledge or ignorance, blood flow increases in a consistent pattern of brain regions: medial prefrontal cortex (BA9), temporal poles bilaterally (BA38), anterior superior temporal sulcus (BA22), and bilateral temporo-parietal junction extending into posterior superior temporal sulcus (BA39/40/22).

Could any or all of these brain regions be a specialized neural substrate for reasoning about beliefs? The reasons to be cautious with results from the false belief task in developmental psychology apply equally to neuroimaging results. First, there is more to solving the false belief task than a concept of belief, and second, there is more to a concept of belief than passing the false belief task (Bloom & German 2000). In addition, “activity” in the standard subtraction methodology of neuroimaging is only as meaningful or specific as the subtracted control condition. Therefore, we propose two basic criteria for a brain region involved in the attribution of beliefs: generality and specificity. First, the candidate region must show increased activity to any stimuli that invite the attribution of beliefs, both true and false. Second, the response must be specific to belief attribution. That is, the candidate region must not show a high response to the presence of a person per se, or during reasoning about nonmental (false) representations. Another important question is whether brain regions involved in belief attribution may be distinct from those that represent other mental state concepts, such as emotion and goal, which emerge earlier in development. This third question is addressed in the “Neuroimaging: Desires, Perceptions, and Emotions” section below.

Few neuroimaging studies have directly tested the neural activity associated with attributing true beliefs. However, a number of studies have included a control condition in which the protagonist’s action is not based on false belief or ignorance, but on true beliefs and perceptions of the situation (Fletcher et al. 1995,

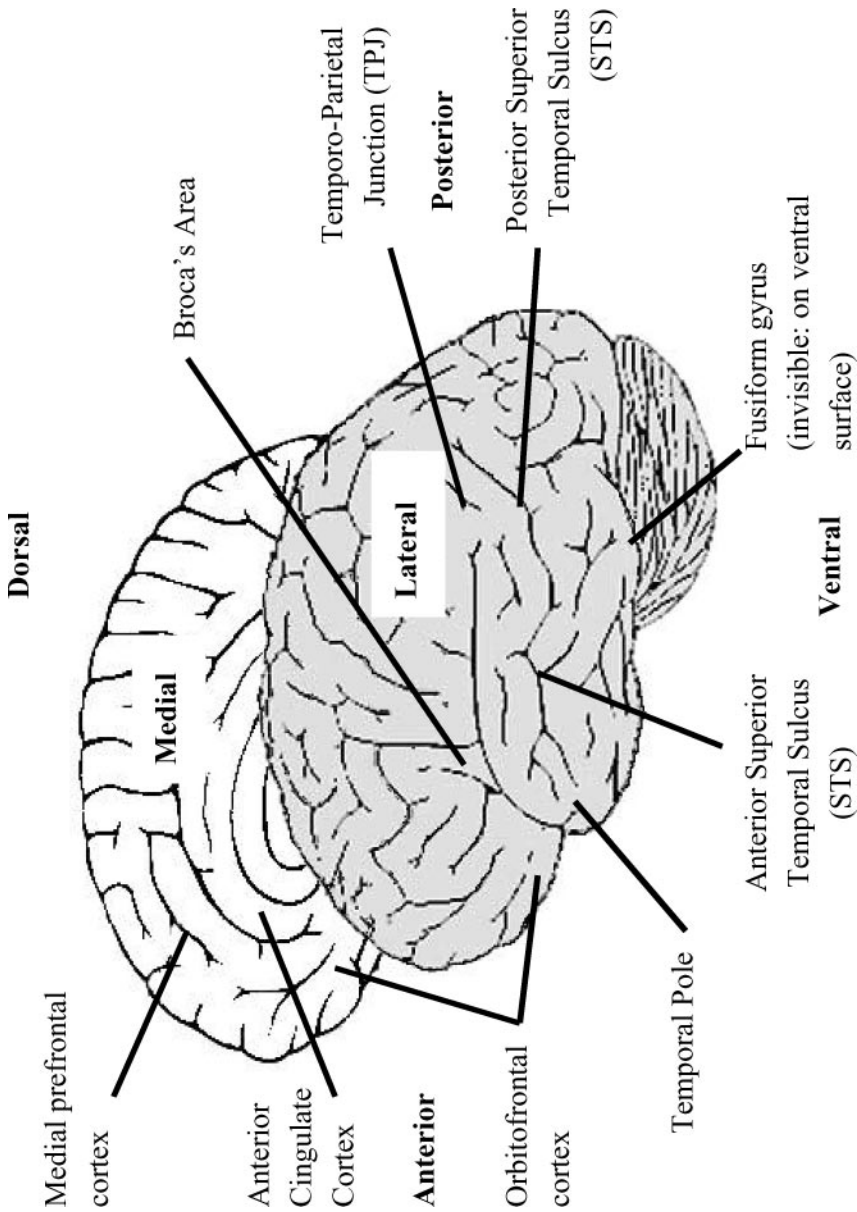


Figure 2 Schematic representation of brain regions associated with the attribution of mental states: beliefs, desires, perceptions, and/or emotions.

Gallagher et al. 2000, Saxe & Kanwisher 2003). This condition may therefore have invited belief attribution, even if reasoning about mental states was not required for successful performance. Consistent with this idea, the same brain regions that showed increased response during false belief stories generalized their activity to the stories that involved true beliefs (Fletcher et al. 1995, Gallagher et al. 2000).² Particularly strong confirmation comes from individual-subject and ROI analyses, showing that the very same voxels in individual subjects are activated for true and false belief attribution (as described in the “Functional Neuroimaging” section above; Saxe & Kanwisher 2003). Thus, the temporo-parietal junction, superior temporal sulcus, and medial prefrontal cortex may all generalize to show a strong activation for both true and false belief attribution.

A candidate region specialized for belief attribution must also be shown to be specific: that is, it must not show a high response during logically similar control conditions that do not require belief attributions. One logical element of reasoning about belief and action is the need to infer invisible causal mechanisms. A second logical ingredient, more specific to false belief stories, which must be included in a control, is the notion of a false representation. Saxe & Kanwisher (2003) therefore had subjects read stories from two control conditions. “Mechanical inference” stories required subjects to infer the operation of an invisible mechanical force (e.g., rusting, evaporation). “False photograph” control stories were modeled on the false photograph paradigm used by developmental psychologists (Leslie & Thaiss 1992, Zaitchik 1990). The false photograph stories can be closely matched to the original false belief stories. In a typical false photograph scenario, a photograph is taken of the scene (e.g., the chocolate in the green cupboard). After the target object has been moved (e.g., to the blue cupboard), the subject must reason about the contrasting states of affairs in the world and in the photograph, analogous to reasoning about a false belief. None of the brain regions that produced a strong response to false belief stories showed a high response to either of these logical control conditions, consistent with the hypothesis that these regions respond specifically during attribution of beliefs, not to any false representation or hidden cause.

Finally, any brain region specialized for representing beliefs must not respond to just the simple presence of a person present in the stimulus. Most studies have

²The only brain region that did not show this pattern was medial frontal cortex, which showed a significantly increased response only during false belief stories and not during stories about actions based on true beliefs (compared with jumbled sentence controls; Fletcher et al. 1995, Gallagher et al. 2000). A number of authors (Gallagher & Frith 2003, Gallagher et al. 2000) have concluded that medial frontal cortex is the only cortical region uniquely involved in “theory of mind.” But, as we and others (e.g., Bloom & German 2000, Scholl & Leslie 2001) have argued, selective involvement in the false belief task per se is not equivalent to selective involvement in “theory of mind.” The medial prefrontal cortex may nevertheless be involved in belief attribution. Saxe & Kanwisher (2003) found that medial prefrontal cortex activity did generalize to vignettes that did not require false belief attribution, and activity in this region has also been reported when subjects are simply instructed to try to understand a character’s motivations (Ferstl & von Cramon 2002).

included a person in a control condition (Fletcher et al. 1995, Gallagher et al. 2000). Even when subjects were forced to attend to the details of a physical description of a person, the responses of the temporo-parietal junction regions bilaterally and the right anterior superior temporal sulcus were no higher than a control condition using physical description of nonhuman objects (Saxe & Kanwisher 2003). Recruitment of these regions requires thinking about a person's beliefs, not just a person's appearance.

Thus a number of brain regions, including bilateral regions of the temporo-parietal junction, posterior and anterior superior temporal cortex, and temporal pole appear to fulfill at least the basic criteria for the neural substrates of attributing beliefs: generality to both true and false belief attribution, and specificity to belief attribution rather than either (*a*) any reasoning about people or (*b*) reasoning about nonmental false representations or hidden causes in general.

Neuroimaging: Belief Attribution and Inhibitory Control

Given the characteristic pattern of brain activation associated with belief attribution, we can begin to address the theoretical questions raised in the first half of this section. First we consider whether brain regions associated with inhibitory control are recruited during (false) belief attribution.

Executive (or inhibitory) control has multiple distinct components, each of which could contribute to false belief attribution, including monitoring and detecting the conflict between competing representations or responses, selecting the correct response, and inhibiting the incorrect (possibly prepotent) response. A recent series of elegant neuroimaging studies have attempted to distinguish the neural correlates of these components (e.g., Botvinick et al. 1999, Braver et al. 2001, Garavan et al. 2002, Konishi et al. 1999, Milham et al. 2001, Sylvester et al. 2003).

Across a range of tasks, including the Eriksen flanker task³ (Botvinick et al. 1999, Eriksen & Eriksen 1974), the Stroop task⁴ (Milham et al. 2001, Stroop 1938),

³In the Eriksen flanker task, subjects make a response determined by the target object in the center of the array. On either side of the target, distractor objects are presented. Distractor objects may be associated with no response (neutral), the same response as the target (congruent), or a different response from the target (incongruent). Botvinick et al. (1999) found that the response of the anterior cingulate cortex was highest on incongruent trials that followed congruent trials, when the selective attention of subjects was relatively relaxed, thus provoking strong response conflict.

⁴In the Stroop task (Stroop 1938) subjects are required to name the ink color of a word. Response conflict increases when the word is the name of a different color (e.g., the subject sees the word "blue" printed in red ink, and must respond "red."). Milham et al. (2001) reported that anterior cingulate cortex activity was maximized when the word named a color in the response set (i.e., a possible ink color, with an assigned response), compared with trials in which the word named a color not in the response set. This finding is consistent with the idea that anterior cingulate cortex activity is an index of response conflict.

and the Go/No-Go task (Braver et al. 2001), response conflict is strongly correlated with activity in anterior cingulate cortex (ACC). Braver et al. (2001) found that a region of ACC is activated for any low-frequency response condition, whether it requires a response (“target detection”) or the suppression of a response (“no-go” trials). This ACC activation is significantly posterior to the paracingulate region of activity associated with false belief problems (e.g., Gallagher et al. 2000). Brain regions consistently associated with response inhibition also include dorsolateral prefrontal cortex (BA 46/9) and superior parietal lobe (BA7; e.g., Braver et al. 2001, Sylvester et al. 2003). Selecting the appropriate response is associated with additional activity in bilateral frontal eye fields and intraparietal sulcus (Jiang & Kanwisher 2003). None of these brain regions are among those associated with reasoning about beliefs.

Thus belief attribution—even of false beliefs—appears to rely on distinct neural systems from those responsible for response conflict, selection, and inhibitory control. Consistent with this conclusion, the stories that describe a nonmental false representation—the false photograph stories—require inhibitory control similar to false belief stories, but do not elicit responses in regions associated with belief attribution (Saxe & Kanwisher 2003). At least for adults, then, false belief attribution may not depend on inhibitory control during task performance. This is consistent with Wellman et al.’s (2001) interpretation of the developmental correlation between inhibitory control and theory of mind: namely, that inhibitory control may help children learn about the mind.

An interesting exception is the target detection paradigm. When subjects are required to respond to a low-frequency, unpredictable, or unexpected target, brain activity increases in the temporo-parietal junction (Braver et al. 2001, Corbetta et al. 2002, Downar et al. 2002, Shulman et al. 2002), in a region near that activated by false belief tasks. Again, however, no study has directly compared these two paradigms within a single experiment. Are there distinct but neighboring subregions of the temporo-parietal junction involved in belief attribution and target detection? Or is there a functional relationship between these two tasks? This issue remains open for future work.

Neuroimaging: Belief Attribution and Language

As described above, de Villiers (2000) has advanced a specific, and testable, hypothesis about the relationship between language and belief attribution: that language is “the only representational system that could” support the concept of a false belief, because it allows the syntactic construction of embedded sentence complements. If so, then we would predict that both belief attribution and sentence-level syntax recruit the same neural structures. Do they?

Studies that vary the syntactic complexity of sentences often find brain activation in and around Broca’s area in left inferior frontal cortex (Caplan 2001). The location of this region of activation is inconsistent across studies, but is not in close proximity to any of the regions implicated above in belief attribution. However, two

other regions are also associated with syntax in some studies: left anterior superior temporal sulcus, near the temporal pole (Ferstl & von Cramon 2002, Friederici 2001, Vandenberghe et al. 2002), and posterior regions of the superior temporal sulcus, near Wernicke's area (Caplan 2001, Ferstl & von Cramon 2002, Just et al. 1996, Roder et al. 2002).

How do the regions of superior temporal sulcus that are involved in syntax compare to the regions of superior temporal sulcus implicated in belief attribution? One study has included versions of the two tasks within a single experiment. Ferstl & von Cramon (2002) scanned subjects while they read sentence pairs in one of two conditions. First, in the logic condition, subjects read two sentences describing a mechanical causal sequence, and were asked to judge whether the sequence was coherent.⁵ In the second half of the experiment, subjects read sentences describing an event involving people, and were asked to try to "understand their feelings, motivations and actions." The controls for both conditions were sentences in pseudolanguage. Compared to this control, the "logic" and the "theory of mind" conditions elicited increased brain activity in strikingly similar regions of posterior and anterior superior temporal sulcus and temporo-parietal junction. Does this mean that syntax and belief attribution recruit similar regions of the posterior superior temporal sulcus (pSTS)? Unfortunately, it is not clear that Ferstl & von Cramon have isolated the parts of superior temporal sulcus responsible for syntax.⁶ In their pseudolanguage control condition, half of the sentences contained familiar syntax, with pseudowords in place of content nouns. Sentences like this, with the syntax but not the content of real sentences, activate the syntax-related region of superior temporal gyrus almost as strongly as normal sentences (Roder et al. 2002). Moreover, the focus of activity associated with syntax in group analyses is typically located 2 or 3 cm anterior to the focus of activity associated with belief attribution).

⁵In their paper, Ferstl & von Cramon (2002) focused on the effect of coherence. In the "logic" condition, the response of the region of medial prefrontal cortex associated with false belief attribution increased significantly only when the pair of sentences was judged to be coherently connected. This may be important because a number of the early studies of false belief tasks (e.g., Fletcher et al. 1995, Gallagher et al. 2000) used random unlinked sentences as their baseline control condition. Ferstl & von Cramon concluded that the medial frontal cortex is responsible generally for "the maintenance of nonautomatic cognitive processes."

⁶It is a serious challenge to explain what did account for the activation of belief attribution-related regions of superior temporal sulcus in the "logic" condition of Ferstl & von Cramon (2002). Frith & Frith (2003) speculate that some of the stimuli may have invited attribution of beliefs and human actions in spite of the absence of people in the explicit descriptions. However, this seems unlikely: The "logic" stimuli were similar to the "mechanical inference" control stories used by Saxe & Kanwisher (2003) that did not elicit any response in the temporo-parietal junction or superior temporal sulcus regions. A different possibility is that the "coherence" instructions used by Ferstl & von Cramon in the "logic" condition encouraged subjects to consider the intentions of the author of the passage.

The relationship between the neural correlates of syntax and of belief attribution remains open to investigation, but the strong conclusion that syntax and belief attribution recruit the same brain regions seems unlikely. Neuropsychological evidence provides an additional hint that these two functions may be independent, at least in adulthood. Siegal and colleagues (Varley & Siegal 2000, Varley et al. 2001) have shown that two dense aphasics with dramatically impaired syntactic processing following strokes, nevertheless can attribute beliefs and even pass a nonverbal false belief task.

Belief Attribution: Conclusions

Developmental psychology and neuroimaging studies provide converging evidence for an anatomically and functionally distinct system for belief attribution in normally developing human children and adults. Children begin to use a novel representational concept of belief around age 3, and by age 4 this concept is robust enough to support successful performance on false belief tasks. The attribution of beliefs also seems to be associated with distinct brain regions, including the medial prefrontal cortex and temporo-parietal junction. One theoretical question about the emergence of belief attribution is the extent to which it depends only on maturation of other capacities. False belief task performance is correlated with both inhibitory control and syntax development. However, we have shown that these capacities recruit different regions of the brain, suggesting that distinct mechanisms are involved in belief attribution, inhibitory control, and syntax. The developmental correlation may therefore reflect the facilitation of knowledge development about the mind for children with mature inhibitory control or linguistic skills.

The most striking dissociation in the development of theory of mind, though, is that between the late development of the concept of belief, described above, and the much earlier development of attribution of other mental states, including desires, perceptions, and emotions (e.g., Bartsch & Wellman 1995). Henry Wellman has characterized this difference as a transition from “desire psychology” to “belief-desire psychology” (e.g., Bartsch & Wellman 1995). Many models of theory of mind include a similar distinction, although the precise characterizations and developmental time courses differ (e.g., Baron-Cohen 1997, Leslie 2000, Tager-Flusberg & Sullivan 2000). In the next section, we will investigate whether a similar distinction is instantiated in the brain regions recruited by theory of mind reasoning.

DEVELOPMENTAL PSYCHOLOGY: DESIRES, PERCEPTIONS, EMOTIONS

Passing the false belief task is sufficient but not necessary evidence of having a theory of mind. According to the criteria originally set out by Premack & Woodruff (1978), young preschoolers and even infants possess a full theory of mind

because they can impute unobservable mental states to themselves and others, and use these mental states as a coherent framework to make predictions about behavior. The central notion in this earlier theory of mind seems to be a concept of desire or goal, but it also includes concepts of perception (or attention) and emotion. In this section, we discuss each of these concepts and consider whether they are processed and represented in one common system or in multiple distinct systems. In each case, we discriminate between two alternative accounts of the children's competence. According to the mentalistic (rich) interpretation, young children attribute mental states to others in order to predict and explain behavior. According to the lean interpretation, children have simply developed (or possess innately) sensitivities to certain physical cues, like changing eye gaze or a smile, without making any mental state attributions (Povinelli 2001). The evidence in this section that children do indeed make coherent, causally interrelated mental state attributions of desires/goals, perceptions, and emotions will help to establish the criteria for the neuroimaging studies that are examined in the "Neuroimaging: Desires, Perceptions, and Emotions" section below.

Desires/Goals

Children make genuine psychological references to desires in their spontaneous speech by their second birthday; references to beliefs appear 6 to 12 months later (Bartsch & Wellman 1995). In the lab, 2-year-old children respond appropriately to another person's desires, even when they differ from the child's own preferences [e.g., giving an adult experimenter more of the snack that she preferred (broccoli) rather than the one the child liked better (crackers) (Repacholi & Gopnik 1997; see also Rieffe et al. 2001)]. Fifteen-month-old children can distinguish between the intended goal of an action and its accidental consequences, selectively imitating the goal (Carpenter et al. 1998a, Meltzoff 1995). Furthermore, young children's attribution of goals is not restricted to human actors, but includes nonhuman agents capable of contingent interactions (Gergely & Csibra 1997, Johnson et al. 2001; but see Legerstee & Barillas 2003).

Strikingly, the concept of a goal seems to be available to preverbal infants. Five- to eight-month-old infants who have habituated to a reach-and-grasp motion by a human hand look longer when the object of the motion changes than when the physical path of the motion changes (Woodward 1998). That is, these infants appear to impute an unobservable mental state to the agent (a desire or goal to have the target object), and use this attribution to make a prediction about future behavior (that the agent will continue to reach for the same object and not, for instance, toward the same location in space).

What is the nature of this early concept of desire/goal? Although the concepts of desire, goal, and intention play distinct roles in adult speech and reasoning (e.g., Malle & Knobe 2001), Astington (2001b) has argued plausibly that toddlers have instead an undifferentiated notion of a volitional state (or "conation") tied to an action or an object. Critically, the toddler does not attribute knowledge or beliefs

about the apple to Anne (“Anne thinks there is an apple”), but simply uses her own knowledge of the world (“There is an apple”) plus a volitional connection from Anne to the apple (approximately “Anne wants the apple”) (Figure 1*b*) to predict Anne’s action (Bartsch & Wellman 1995, Wellman & Cross 2001).⁷ An undifferentiated concept of volition is in principle sufficient to support everything from the 5-month-old’s simple goal-directed interpretation of reaching (Woodward 1998) to the 24-month-old’s understanding that different people can have different preferences (Repacholi & Gopnik 1997). However, it remains an open question whether the concept of desire/goal is essentially unchanged in the first two years of life and merely becomes more robust, or whether there is real conceptual change from a representation of a goal of hand actions in particular to a more general notion of desire.

Thus young children can attribute desires and goals long before they attribute beliefs. However, an important criterion of a theory of mind is that different mental state attributions interact in a coherent causal framework in order to allow explanations and predictions of action. In the next two subsections, we present evidence for such interactions between the early concepts of desire/goal, perception, and emotion.

PERCEPTIONS Like the concept of desires/goals, a preliminary concept of perception (i.e., another person’s ability to see or look at something) is available to children long before belief attribution. Verbs for seeing are in children’s productive vocabulary by 26 months (Bretherton & Beeghly 1982), and children spontaneously produce embedded-clause syntax with “see” six months earlier than with “think” (Bartsch & Wellman 1995). The earliest sensitivity to others’ eyes is evident before infants are 3 months old, expressed by a preference for faces with open eyes (Bakti et al. 2000) and by orientation toward the direction of gaze of a previously viewed face (Hood et al. 1998). However, early (and later) behavioral sensitivity to another’s gaze is open to both rich and lean interpretations. According to the rich interpretation, infants (or young children) understand seeing as a mental state, providing the actor with selective visual access to (and possibly knowledge about) the world. According to a lean interpretation, on the other hand, gaze following does not demonstrate any mentalistic understanding (see Povinelli 2001); rather, the tendency of infants to follow gaze could depend on a learned association between the direction of adult gaze and interesting events in the world or on a hard-wired reflex (e.g., Corkum & Moore 1995).

⁷This characterization helps to resolve the puzzle of how the concept of desire could precede, and be independent of, the concept of belief. In some frameworks (e.g., Searle 1983) the concepts of desire and belief are logically identical, each composed of a proposition (“It will rain tomorrow”) and an attitude (either “I want” or “I think”). However, according to this interpretation, young children do not have such a representational concept of desire. Rather, they may conceive of a desire as a direct connection between a person and a real object. If so, a toddler could understand desires without necessarily understanding representations.

When do children possess a genuinely psychological concept of perception? By early in their second year, children's concept of perception shows two hallmarks of being incorporated in their theory of mind: perception is understood to be referential, and to interact coherently with other attributed mental states, including goals and emotions. (We discuss this second point in the section on emotion understanding.)

A referential concept of perception specifies that gaze is directed *at* something; i.e., that looking is a relationship between a person and an object. A series of recent studies show that by 12 to 14 months, infants interpret others' gaze as referential (Brooks & Meltzoff 2002, Caron et al. 2002; but see Doherty & Anderson 1999). Brooks et al. (described in Caron et al. 2002) adapted Woodward's paradigm from goal-directed actions to measure referential understanding of perception. Fourteen-month-old infants were habituated to an adult with either open or closed eyes, who turned toward one of two objects. When the position of the objects was switched at test, the infants looked longer if the adult now looked to the old location (new object) than if the adult looked to the new location (old object), but only if the adult's eyes were open during habituation, suggesting that the infants understood that looking, but not head-turning, is object-directed. Gaze following is also not restricted to human gaze. Johnson et al. (1998) found that 12-month-old infants would follow the "gaze" of a faceless unfamiliar object after only 60 seconds of contingent interaction between the baby and the object, a finding that suggests gaze following at this age is not merely a conditioned response.

Infants use their mentalistic understanding of gaze in word learning. Dare Baldwin (1993) gave 14- and 18-month-old infants one object to play with, while another object was put into a bucket in front of the experimenter. When the infant was looking at the object she was playing with, the experimenter looked into the bucket, and introduced a novel noun: "It's a blicket!" This introduces a perfect perceptual association between the novel word and the object in the infant's hand. Instead of learning this association, the infants looked up to see where the experimenter was looking. The 1-year-olds then associated the new word "blicket" with the object that the experimenter was looking at, and not the one in their hand.

While the early concept of perception is referential, it is probably not representational. As with goals, young children may conceive of another person's perception as a direct connection between the person and the real object in the world (without positing any internal mental representation, Figure 1). This distinction helps to explain the long delay from the (referential) understanding that perception is object-directed, around 14 months, to the (representational) notion that perception can be inaccurate or only partial and so can lead to misperceptions, which is still developing in children 4 and 5 years old (e.g., Gopnik & Astington 1988, Lalonde & Chandler 2002).

EMOTIONS Newborn infants already show sensitivity to, and synchrony with, others' emotions: they cry when other infants cry, and show some discrimination of happy versus sad expressions (Field & Walden 1982). By 6 months, infants

undeniably discriminate facial expressions (Caron et al. 1988, Nelson 1987), and by their first birthday, infants use parental emotional expressions, for example of fear, to guide their own actions in novel situations (“social referencing,” Feinman 1992). But the discrimination of emotional expressions is open to a lean, nonmentalistic interpretation. For instance, the physical configuration of emotional expressions may constitute (either learned or innate) intrinsic rewards or punishments to the perceiver (e.g., Blair 2003). When, then, do the emotions of others become incorporated in infants’ emerging theory of mind? As mentioned above, one hallmark of such incorporation is the ability to form coherent combinations of attributed mental states, which emerges around age 14 months.

By 14 months, infants can combine information about a person’s gaze and emotion both to infer the person’s goal and to direct the infant’s own actions. Phillips et al. (2002) habituated infants to an event in which an adult looked and smiled at one of two available objects, and then was shown holding the same object. At test, the adult looked and smiled either at the same (old) object or at the other (new) object, and then was shown holding the new object. Infants looked longer when the adult took the new object after gazing at the old object, which suggests that infants used gaze and emotion cues to infer the actor’s subsequent goal.⁸ Infants can also use emotional expressions to infer a previous goal. In a study by Tomasello et al. (1996), an adult used a nonword to announce an intention to find an object. The adult first picked up one object with obvious disappointment and rejected it, and then picked up a second object with glee (all nonverbally). Sixteen-month-old infants learned that the novel word referred to the object of the positive emotion, which suggests that infants know that happiness results from goal fulfillment. Finally Moses et al. (2001) adapted Baldwin’s (1993) word learning paradigm to show that infants use the gaze direction of an adult to determine the referent of an emotional message. When an adult made a negative emotional noise, 12- and 18-month-olds looked up from the novel object in their hand to determine the gaze direction of the adult. Infants subsequently avoided only the object that had been the focus of the adult’s gaze.

The causal relation between attributions of emotions and desires is also apparent once children begin to speak. Around 24 months, children begin to spontaneously talk about the causes of their own emotions (Bretherton et al. 1986, Wellman et al. 1995), especially the relationship between frustrated desires and negative mental states (Dunn & Brown 2001, Lagattuta & Wellman 2001). In the lab, 2-year-olds who were told about a boy who wanted a puppy and got one choose a happy face to show how the boy would feel, but the same children choose a sad face if the boy had wanted a bunny (Wellman & Woolley 1990).

⁸In an elegant control, the authors showed that if the contingency was reversed during habituation (i.e., the adult looked at one object, and then was shown with the other one), infants were not able to generalize this pattern to the test trials. Thus, infants were not simply learning a pattern of contingencies during the experiment.

Of course, as with concepts of perception and desire, children's understanding of emotions continues to develop after age 2. Complex social emotions, like pride, embarrassment, and guilt begin to be correctly attributed between the ages of 5 and 14 (e.g., Berti et al. 2000). But the framework for attributing basic emotions and their causal relations with desires and perception appears to be already intact in the second year of life.

COMMON OR DISTINCT MECHANISMS? In all, in their second and third year, toddlers reason productively and coherently about action, using basic concepts of (and causal relations between) three kinds of mental states: desires or goals, perceptions, and emotions. This early theory of mind seems to emerge significantly earlier than the reasoning about beliefs described in the first half of this review.

An open question concerns the extent to which attributions of desires, perceptions, and emotions rely on distinct or common functional or anatomical substrates. For instance, Alan Leslie (1994), in his hierarchical model of theory of mind development, lumps perceptions and goals together as actional properties represented by a stage he calls the Theory of Mind Module 1. Actional properties are those that let an agent "act in pursuit of goals, react to the environment and interact with each other." Consistent with this proposal, in a longitudinal study of 9- to 15-month-olds, Carpenter et al. (1998b) found that the emergence of attentional engagement and gaze following (attribution of perception) and imitation of novel actions (goal attribution) were positively correlated with each other, and uncorrelated with concurrent nonsocial developments such as object permanence. The authors conclude that a mentalistic understanding of gaze and action are "two instances of the same underlying phenomenon." Simon Baron-Cohen (1994, 1997), on the other hand, has divided this domain into two distinct components. The "Intentionality Detector" represents behavior in terms of goals, while the "Eye Direction Detector" detects eyes and represents the direction as the agent "seeing." Neither model explicitly includes a mechanism for emotion attribution.

Neuroimaging: Desires, Perceptions, and Emotions

The above review suggests three central questions that could be addressed using neuroimaging. First, reasoning about beliefs develops later than, and may depend upon (e.g., Pellicano & Rhodes 2003), an earlier theory of mind that includes attribution of desires, perceptions, and emotion. Does the later emerging competence colonize the same neural systems that underpin earlier reasoning? If so, we would predict that attributions of desires, for instance, would recruit the same brain regions identified above as involved in belief attribution. If, on the other hand, reasoning about beliefs draws on distinct systems or abilities, then desire attribution should not produce activity in regions associated with belief attribution, and may recruit a distinct set of brain regions.

Second, the models of early theory of mind proposed by Leslie (1994) and Baron-Cohen (1994, 1997) disagree about whether attributions of perception and

goals are the province of one common or two distinct modules. A preliminary approach to resolving this controversy is to ask whether these functions recruit the same or different brain regions in adults.

Finally, what is the relationship between the attribution of emotions and of other mental states?

NEUROIMAGING: ATTRIBUTING DESIRES AND GOALS In functional neuroimaging studies, the attribution of desires, goals, and intentions to another person has been investigated in three basic paradigms. Studies of the first kind invite mental state attribution to a fictional character using vignettes, cartoons, or animations similar to the stimuli used in studies of belief attribution (Brunet et al. 2000; Castelli et al. 2000, 2002; Fletcher et al. 1995; Gallagher et al. 2000; Saxe & Kanwisher 2003; Schultz et al. 2003). In a second, related set of studies, subjects engage in a simple game, purportedly either with an unseen agent (presumably inviting goal attributions) or with a computer (discouraging goal attributions) (Gallagher et al. 2002, McCabe et al. 2001). Finally, in the third kind of study, subjects watch and interpret a video of a simple goal-directed action by a human actor (Chaminade et al. 2002; Decety et al. 2002; Koski et al. 2002; Zacks et al. 2001; R. Saxe, D.K. Xiao, G. Kovacs, D. Perrett, & N. Kanwisher, submitted). Below we consider the three sets of studies sequentially. Unfortunately, few of these studies also included a task designed to elicit belief attribution, and to our knowledge only one study to date explicitly aimed to contrast brain regions involved in the attribution of different kinds of mental states (R. Saxe, D.K. Xiao, G. Kovacs, D. Perrett, & N. Kanwisher, submitted).

Vignettes, cartoons, and animations that depict or suggest a character's goals, intentions, or desires are typically correlated with moderately increased activity (compared to scrambled or nonsocial controls) in the brain regions associated with belief attributions, including medial prefrontal cortex and posterior superior temporal sulcus (Brunet et al. 2000, Buccino et al. 2001, Castelli et al. 2000, Gallagher et al. 2000, Saxe & Kanwisher 2003, Schultz et al. 2003). For instance, stories describing a character's desires elicited significantly more activity in these regions than physical descriptions of a person, but significantly less than stories describing false beliefs (Saxe & Kanwisher 2003). The most important weakness of this evidence is that none of the stimuli in these studies were designed to exclude belief attribution. The intermediate activity in all of these studies may reflect weak but consistent activation of these regions during desire/goal attribution, but it may equally reflect subjects' occasional spontaneous belief attribution in response to these stimuli. Future studies are needed in which vignettes about different mental states are explicitly contrasted.

Games provide an appealing paradigm for investigations of goal attribution because they allow the stimuli to be exactly matched during goal-attribution and no-goal conditions. Two PET studies using this logic found increased activity in regions of medial prefrontal cortex when the opponent was (purportedly) a human, compared with computer-opponent trials (Gallagher et al. 2002, McCabe et al.

2001). This contrast is too broad to isolate goal attribution in particular, though. Differences in brain activity could reflect belief attribution as above, or just the felt presence of a human opponent [see description in Gallagher et al. 2002; Saxe & Kanwisher (2003) reported that a region of medial prefrontal cortex responded more to any story containing a person, including physical descriptions, than to nonhuman control stories]. To date, no studies have directly compared belief and goal attribution within the context of a game.

Many studies have investigated the neural correlates of observing human body movements. Here we concentrate on the subset of these studies that explicitly address the perception of intentional or goal-directed action. One context in which representation of goal-directed action has been investigated is imitation (Chaminade et al. 2002, Decety et al. 2002, Koski et al. 2002). For example, Koski et al. (2002) asked subjects to imitate simple index finger movements, viewed either with or without target dots (the target dots presumably made the finger movement appear goal-directed). The presence of the goal produced increased activation in lateral inferior frontal cortex bilaterally (Broca's area, also observed by Buccino et al. 2001 for object-directed versus mimed movements of hands and mouths). Furthermore, transient disruption of Broca's area by transcranial magnetic stimulation (TMS) interfered with imitation, but not with cued execution, of goal-directed finger movements (Heiser et al. 2003). Broca's area is not one of the regions associated with belief attribution, which provides preliminary evidence that at least very simple goals may be attributed using different brain regions from those involved in reasoning about beliefs.

Finally, two studies have investigated brain regions associated with the segmentation or interpretation of whole body actions. Zacks et al. (2001) looked for activity correlated with event boundaries in, or transitions between subgoals of, a complex goal-directed action (e.g., cleaning the kitchen). Saxe et al. (R. Saxe, D.K. Xiao, G. Kovacs, D. Perrett, & N. Kanwisher, submitted) varied the structure of a simple intentional action (walking across a room) by directing the actor to unexpectedly pause for a few seconds behind a large bookcase, perhaps requiring subjects to reformulate their interpretation of the action (the occlusion manipulation allowed the two conditions to be matched for average visual information, including biological motion). Both studies report activity related to action segmentation in right posterior superior temporal sulcus. Saxe et al. (R. Saxe, D.K. Xiao, G. Kovacs, D. Perrett, & N. Kanwisher, submitted) further established that this activation was specific to intentional actions, because the same pattern was not observed for passively occluded people, and generalized to other stimuli, because the same region showed a high response to two-dimensional animations portraying goal-directed actions compared with rapid rigid rotation. Activity in the same vicinity was also reported by Decety et al. (2002) when subjects viewed another person's hands performing an action similar to the action the subjects were concurrently executing, compared with viewing their own hands performing that action.

Is the region of posterior superior temporal sulcus that is involved in the analysis of intentional action the same as the nearby region associated with the attribution

of beliefs? Saxe et al. (R. Saxe, D.K. Xiao, G. Kovacs, D. Perrett, & N. Kanwisher, submitted) argue that these two regions are distinct. The regions showing increased activity during false belief stories (dubbed the temporo-parietal junction region) and during the paused walking action (dubbed the pSTS visual analysis of action) did not overlap anatomically in individual subjects, and showed strikingly different functional profiles in ROI analyses.

The evidence is inconclusive as to whether interpreting the goal or intention of an observed action draws on distinct brain regions from those involved in assigning beliefs to the actor. Evidence for distinct brain regions for attributing goals comes from studies using videos of simple actions (e.g., Heiser et al. 2003; Koski et al. 2002; R. Saxe, D.K. Xiao, G. Kovacs, D. Perrett, & N. Kanwisher, submitted). Weaker evidence for common regions recruited during both belief and desire attributions comes from studies using vignettes, cartoons, animations, and games, although all of these studies have tended to confound attributions of beliefs and desires (Brunet et al. 2000; Castelli et al. 2000, 2002; Fletcher et al. 1995; Gallagher et al. 2000, 2002; McCabe et al. 2001; Saxe & Kanwisher 2003). One possible resolution is that the brain contains distinct representations of goals (restricted to simple, visible motor actions), similar to the notion of goal available to 5- to 8-month-olds, and desires (applicable more generally), available after the first birthday. This speculation awaits testing in future work.

To summarize, brain regions involved in representing goal-directed action (including pSTS and Broca's area) are distinct from the brain regions associated with belief attribution (including the temporo-parietal junction and medial prefrontal cortex). Thus brain activation patterns are consistent with developmental psychology in suggesting distinct mechanisms for attributing goals and beliefs. As summarized above, an understanding of goal-directed action is available to even very young infants, while belief attribution does not emerge until three years later. The critical, and unresolved, question remains the place of desires in this scheme. Young toddlers talk about desires long before they talk about beliefs (Bartsch & Wellman 1995), and seem to conceive of desire and goals similarly, as direct volitional connections to the world (Astington 2001a,b), which suggests that desire attribution should be similar to goal attribution and distinct from belief attribution. For adults, on the other hand, belief and desire attribution may be simply inseparable. In neuroimaging studies using vignettes, cartoons, animations, and games, attributions of desires and beliefs seem to elicit activity in the same set of regions [although beliefs and desires were always confounded in the stimuli (Brunet et al. 2000; Castelli et al. 2000, 2002; Fletcher et al. 1995; Gallagher et al. 2000, 2002; McCabe et al. 2001; Saxe & Kanwisher 2003)]. Alternatively, desires may be reanalyzed as representational as part of the conceptual transition to a representational treatment of belief. In adult folk theory, we desire a particular apple under a particular description, e.g., representing it as food that tastes good. Further work is necessary to determine whether the correct division of "natural kinds" in the adult brain places desires with goals or with beliefs.

NEUROIMAGING: ATTRIBUTING PERCEPTION In developmental psychology, a controversy remains over whether the attribution of goals and of perceptions relies on a single system (e.g., Leslie 1994), or two distinct systems (e.g., Baron-Cohen 1997). One way to address this question is to ask whether attributions of goals and of perception recruit the same or distinct brain regions in adults.

Activity in the right pSTS is associated with perception of the gaze of a face: The response in this region is higher for moving eyes than for a moving checkerboard (Puce et al. 1998) or for a change in facial identity (Haxby et al. 2002), for open (direct or averted) eyes than for closed eyes (Wicker et al. 1998), and when subjects attend to the gaze rather than the identity of faces (Hoffman & Haxby 2000). However, as described above for the developmental studies, a sensitivity to gaze is not sufficient to indicate the attribution of perception. Perception is referential, gaze-directed toward something. In an elegant recent study, Pelphrey et al. (2003) showed that the response of the right pSTS to moving eyes is modulated by interaction with a target—a small checkerboard to the right or the left of the character's face. When the character quickly moved his eyes away from the target (incongruent) instead of toward the target (congruent), the response of the pSTS was sustained for many seconds longer. Pelphrey et al. (2003) suggest that on incongruent trials “the observer's expectation is violated and activity in the STS region is prolonged—perhaps related to a reformulation of an expectation.”

What is the relationship between regions of the pSTS⁹ associated with perception of gaze changes and of other intentional actions? To our knowledge, no single study has combined target-directed hand or body actions and gaze changes. Puce et al. (1998) reported that eye and mouth movements (with no goal) elicited activity in the same part of pSTS, but other studies using hand and body actions appeared to produce activity more lateral and anterior in the STS. On the other hand, centers of activity reported by Pelphrey et al. (2003) for unexpected gaze change and by Saxe et al. (R. Saxe, D.K. Xiao, G. Kovacs, D. Perrett, & N. Kanwisher, submitted) for unexpected action changes are anatomically and conceptually similar. A further suggestion of combined neural representation comes from the discovery of neurones in the anterior superior temporal sulcus of macaque monkeys that show increased response to target-directed hand actions only when the actor's gaze is directed toward the action (Jellema et al. 2000). Neuroimaging work along these lines may help to determine whether attributions of referential gaze and goal-directed actions rely on common or distinct brain regions, consistent with Carpenter et al's (1998b) conclusion described above that these two behaviors reflect “the same underlying phenomenon.”

⁹Eye gaze—especially averted gaze—was also associated with activity in medial prefrontal cortex, compared with eyes looking down or closed (Calder et al. 2002). These authors point out the anatomical similarity between this activation and the region of medial prefrontal cortex associated with false belief task performance (e.g., Gallagher et al. 2000), and suggest that gaze information may recruit activity across the whole network of brain regions associated with theory of mind.

NEUROIMAGING: ATTRIBUTING EMOTIONS Investigations of emotion have tended to proceed separately from investigations of theory of mind (but see Terwogt & Stegge 1998). Nevertheless, emotions can be attributed to others and are causally interrelated with other mental state attributions: fulfilled goals cause happiness, the object of fear can be identified by gaze direction, etc.

An overview of the neural systems associated with the perception, experience, and function of emotion is available from many recent reviews (Adolphs 2002, 2003; Blair 2003; Canli & Amin 2002; Cardinal et al. 2002; Hamann 2003; Haxby et al. 2000, 2002; LeDoux 2000; Morris 2002; Preston & de Waal 2002). Facial emotional expressions (usually compared with neutral faces) are associated with activity in a number of different brain regions, including extrastriate cortex, right parietal cortex, right fusiform gyrus, orbitofrontal cortices, amygdala, insula, and basal ganglia (Adolphs 2002).

For the purposes of this paper, we address the narrower question of brain regions recruited during the attribution of emotion to another person and the integration of emotions in a theory of mind. Thus we distinguish neural responses to facial expressions that reflect emotion attribution, from those that reflect, for example, threat detection (e.g., Adolphs & Tranel 2000), the intrinsic reward value of a facial expression (e.g., Blair 2003), resolution of environmental ambiguity (e.g., Whalen 1999), or the initiation of behavioral withdrawal (e.g., Anderson et al. 2000)—all of which may be independently correlated with the perception of facial expressions. We may also distinguish the attribution of an emotion (“she is feeling sad”) from simple emotional contagion (“this makes me feel sad”), although it is controversial whether these are indeed distinct.

One way to look for brain regions involved in the attribution of emotion may be to use descriptions of personal emotional experiences. Along these lines, Decety & Chaminade (2003) asked subjects to watch videos in which actors recounted experiences (in the first person) that were either sad or neutral in content. The actors’ facial and emotional expressions were also manipulated to be happy, sad or neutral. The sad narrative content, irrespective of emotional expression, led to increased neural activity in regions associated with negative emotions (e.g., the amygdala) and in regions associated with belief attribution (e.g., the temporal pole). Interestingly, attribution of emotion was also associated with activity in left lateral inferior frontal cortex (near Broca’s area), which was associated with the representation of goals in the imitation paradigm (Koski et al. 2002). Two distinct functional patterns were observed. In the anterior part of this region (the pars orbitalis) the neural response was high during stories with sad negative content, regardless of the actor’s expression. However, the more dorsal part (the pars opercularis) showed a high response to emotional expression (happy or sad) independent of narrative content. This pattern of results is suggestive of a dissociation between perception of facial emotion and understanding of emotional content in speech. The relationship between representations of goals and of emotions in lateral inferior cortex requires further investigation.

A second way to identify the neural correlates of emotion attribution may be to look for interactions between attributions of emotions and of other mental states.

Wicker et al. (2003) investigated whether specific brain regions are sensitive to the interaction between emotional expression and gaze (specifically, direct versus averted gaze). They reported that the response of a single region in the right anterior superior temporal gyrus was highest during emotional expressions directed at the subject, compared with direct neutral gaze, and the interaction with averted gaze was significant. However, these results are somewhat hard to interpret, since this region has not been previously associated either with the perception of gaze, or with the attribution of emotion. It is not possible to compare the brain regions recruited during attribution of perception and emotion in this study because all conditions included the same gaze shift. Again, this provides no clear evidence that either common or distinct brain regions are involved in the attribution of emotion and perception.

Finally, we might ask whether brain regions identified previously in this review as associated with the attribution of beliefs, goals, or perception are also associated with perception of facial emotion. One such candidate region is the posterior STS.¹⁰ Regions of posterior STS have been associated with representations of action (Decety et al. 2002; R. Saxe, D.K. Xiao, G. Kovacs, D. Perrett, & N. Kanwisher, submitted) and of perception (Pelphrey et al. 2003). Both of these studies found increased activity in the right posterior STS when subjects' expectations about human behavior were violated. Decety & Chaminade (2003) found that a nearby region produced a high response when an actor recounted a negative personal experience using positive emotional facial and vocal expression. Such incongruence between narrative content and affect may constitute a violation of expectations about behavior in the emotional domain. (For further evidence of an association between posterior STS and emotion attribution, see Narumoto et al. 2001). Future work should aim to determine whether these regions of the STS reflect a common neural mechanism for the attribution of goals, perceptions,

¹⁰Another region worth considering in this context is the right fusiform gyrus. A number of studies have reported modulation of the right fusiform gyrus by facial emotional expression (greater response to emotional than neutral faces, e.g., Vuilleumier et al. 2002, Halgren et al. 2000). Geday et al. (2003) found that posterior right fusiform gyrus activity was greater for pictures of emotional complex social scenes than for neutral counterparts. These stimuli were specifically designed to be oriented away from the observer, in order to limit the possibility of direct threat or reward to the subject. Geday et al. therefore claim that their task selectively recruited regions involved in attributing emotions to others. The right fusiform gyrus has also been implicated when subjects view animations of social interactions designed to elicit mental state attributions (e.g., seduction, bullying: Castelli et al. 2000, Schultz et al. 2003). However, it is unclear whether fusiform gyrus activity associated with the animations reflects attributions of other mental states like desires or perceptions, or simply of emotions, since these were confounded in both studies. Further research is needed to determine the relationship between regions of the fusiform gyrus involved in the attribution of emotions and the subregion of the fusiform gyrus (cf. Schultz et al. 2003) known as the fusiform face area (Kanwisher et al. 1997), which shows a greater response for faces than all other familiar object classes.

and emotions, or whether neighboring but distinct subregions are independently responsible for each of these functions.

The question of the relationship between the attribution of emotion and other components of theory of mind remains unanswered. More work is needed on both psychological and anatomical commonalities between these two critical components of understanding others.

SUMMARY: DESIRES, PERCEPTIONS, AND EMOTIONS Evidence from developmental psychology unequivocally supports distinct psychological mechanisms for attributing desires/goals, perceptions, and emotions to others (the early-developing theory of mind) from those mechanisms responsible for attributing beliefs. The results of the neuroimaging studies reviewed above suggest a similar division between brain regions. Videos of simple goal-directed action elicit activity in a region of posterior STS that is distinct from the nearby temporo-parietal junction region associated with belief attribution (Decety et al. 2002; R. Saxe, D.K. Xiao, G. Kovacs, D. Perrett, & N. Kanwisher, submitted). A similar region of posterior STS has also been implicated in attributing perception (Pelphrey et al. 2003) and even possibly emotion (Decety & Chaminade 2003, Narumoto et al. 2001). Future studies are necessary to determine whether this might indeed reflect a single underlying mechanism for representing all of the so-called actional properties of people that let them act in pursuit of goals, react to the environment, and interact with each other (Leslie 1994).

CONCLUSIONS

Substantial behavioral evidence indicates that understanding other minds follows a characteristic developmental trajectory, beginning with the early appearance (in the first 2 years of life) of a system for reasoning about other people's goals, perceptions, and emotions, and the later development (around 4 years of age) of a system for representing the contents of other people's beliefs. Here we asked whether neuroimaging research in adults has contributed or can contribute to theoretical debates about theory of mind that have arisen from the developmental literature. We argue that in several instances, the neuroimaging literature already provides important constraints on these debates.

First, neuroimaging has identified brain regions that are selectively engaged when people reason about the contents of other people's beliefs. This finding strengthens arguments that theory of mind constitutes a special domain of cognition, with its own domain-specific processing machinery. Second, the brain regions associated with belief attribution appear to be distinct from other regions engaged when people reason about other people's goals, which suggests that the two stages of development result from the appearance of two distinct mechanisms, rather than from the gradual enrichment of a single mechanism. Third, the brain regions associated with belief attribution appear to be distinct from those engaged in inhibitory

control and from those engaged in syntactic processing. This finding argues against the hypothesis that these other functions are necessarily engaged when attributing beliefs. Instead, the neuroimaging data suggest that the reported correlations between the development of theory of mind and both inhibitory control and syntactic processing may reflect the requirement of these systems for learning about beliefs.

While these contributions from neuroimaging are substantial, they leave many other important questions unresolved. First, if we can engage reasoning about desires without engaging reasoning about beliefs, will we still see activation of brain regions associated with belief attribution, which would suggest the existence of common mechanisms, or will we fail to engage the same regions, consistent with a real dissociation between reasoning about beliefs and desires? Second, are the early-developing abilities to understand other people's goals, perceptions, and actions based on a single system, or several distinct systems? Third, when strict individual-subjects analyses are applied as described in the "Functional Neuroimaging: Standards of Evidence and Inference" section, are distinct but neighboring subregions of the temporo-parietal junction involved in belief attribution and target detection? If instead these tasks engage overlapping regions, what common process might explain that overlap? Finally, we are hopeful that neuroimaging can also address other important aspects of understanding other minds that we have not had space to address here, such as the perception and neural representation of agency (Csibra 2003; Farrer & Frith 2002; Farrer et al. 2003; Johnson 2000, 2003; Ruby & Decety 2001; Tremoulet & Feldman 2000), the relationship between action perception and action planning (Rizzolatti et al. 2001, Wolpert et al. 2003), the related problem of whether theory of mind is implemented as a pseudoscientific theory or as a simulation (Bartsch 2002, Gallese & Goldman 1998, Nichols et al. 1996, Stich & Nichols 1998), the relationship between moral cognition and emotion (Greene & Haidt 2003, Moll et al. 2002), and the relationship between attribution of enduring traits (like personality) and transient states of a person (like emotions and goals, e.g., Winston et al. 2002; A.S. Heberlein, R. Adolphs, D. Tranel, & H. Damasio, submitted).

In sum, we are optimistic that neuroimaging data can help to answer fundamental questions emerging from developmental psychology about our system for reasoning about other people. These contributions are clearest for questions about the basic architecture of the system for understanding other minds: What are its fundamental components? However, as argued in detail in the "Functional Neuroimaging: Standards of Evidence and Inference" section, neuroimaging can make a real contribution toward answering these questions only if we uphold strict standards concerning the way the data are analyzed and the kinds of inferences we draw from them.

ACKNOWLEDGMENTS

This work was funded by grant NIHM 66,696. We are grateful to Yuhong Jiang, Chris Baker, and Andrea Heberlein for comments and suggestions.

The *Annual Review of Psychology* is online at <http://psych.annualreviews.org>

LITERATURE CITED

- Adolphs R. 2001. The neurobiology of social cognition. *Curr. Opin. Neurobiol.* 11:231–39
- Adolphs R. 2002. Neural systems for recognizing emotion. *Curr. Opin. Neurobiol.* 12:169–77
- Adolphs R. 2003. Cognitive neuroscience of human social behaviour. *Nat. Rev. Neurosci.* 4:165–78
- Adolphs R, Tranel D. 2000. Emotion recognition and the human amygdala. In *The Amygdala: A Functional Analysis*, ed. JP Aggleton, pp. 587–630. London/New York: Oxford Univ. Press
- Allison T, Puce A, McCarthy G. 2000. Social perception from visual cues: role of the STS region. *Trends Cogn. Sci.* 4:267–78
- Anderson AK, Spencer DD, Fulbright RK, Phelps EA. 2000. Contribution of the anteromedial temporal lobes to the evaluation of facial emotion. *Neuropsychology* 14:526–36
- Apperly IA, Robinson EJ. 2002. Five year olds' handling of reference and description in the domains of language and mental representation. *J. Exp. Child Psychol.* 83:53–75
- Astington JW. 2001a. The future of theory-of-mind research: understanding motivational states, the role of language, and real-world consequences. *Child Dev.* 72:685–87
- Astington JW. 2001b. The paradox of intention: assessing children's metarepresentational understanding. See Malle et al. 2001, pp. 85–103
- Astington JW, Jenkins JM. 1995. Theory of mind development and social understanding. *Cogn. Emot.* 9:151–65
- Astington JW, Jenkins JM. 1999. A longitudinal study of the relation between language and theory of mind development. *Dev. Psychol.* 35:1311–20
- Bakti A, Baron-Cohen S, Wheelwright A, Connellan J, Ahluwalia J. 2000. Is there an innate gaze module? Evidence from human neonates. *Infant Behav. Dev.* 23:223–29
- Baldwin DA. 1993. Infants' ability to consult the speaker for clues to word reference. *J. Child Lang.* 20:395–418
- Baldwin DA, Baird JA. 2001. Discerning intentions in dynamic human action. *Trends Cogn. Sci.* 5:171–78
- Baron-Cohen S. 1994. How to build a baby that can read minds: cognitive mechanisms in mindreading. *Cah. Psychol. Cogn. Curr. Psychol. Cogn.* 13:513–52
- Baron-Cohen S. 1997. *Mindblindness: An Essay on Autism and Theory of Mind*. Cambridge, MA: MIT Press. 171 pp.
- Baron-Cohen S, Leslie A, Frith U. 1985. Does the autistic child have a theory of mind? *Cognition* 21:37–46
- Baron-Cohen S, Tager-Flusberg H, Cohen DJ, eds. 2000. *Understanding Other Minds*. London/New York: Oxford Univ. Press
- Bartsch K. 2002. The role of experience in children's developing folk epistemology: review and analysis from the theory-theory perspective. *New Ideas Psychol.* 20:145–61
- Bartsch K, Wellman HM. 1995. *Children Talk About the Mind*. London/New York: Oxford Univ. Press
- Berti AE, Garattoni C, Venturini B. 2000. The understanding of sadness, guilt, and shame in 5-, 7-, and 9-year-old children. *Genet. Soc. Gen. Psychol. Monogr.* 126: 293–318
- Blair RJ. 2003. Facial expressions, their communicatory functions and neuro-cognitive substrates. *Philos. Trans. R. Soc. London Ser. B* 358:561–72
- Blakemore SJ, Decety J. 2001. From the perception of action to the understanding of intention. *Nat. Rev. Neurosci.* 2:561–67
- Bloom P. 2000. *How Children Learn the Meanings of Words*. Cambridge, MA: MIT Press. 300 pp.
- Bloom P, German TP. 2000. Two reasons to abandon the false belief task as a test of theory of mind. *Cognition* 77:B25–31
- Born P, Leth H, Miranda MJ, Rostrup E, Stensgaard A, et al. 1998. Visual activation in

- infants and young children studied by functional magnetic resonance imaging. *Pediatr. Res.* 44:578–83
- Botvinick M, Nystrom LE, Fissell K, Carter CS, Cohen JD. 1999. Conflict monitoring versus selection-for-action in anterior cingulate cortex. *Nature* 402:179–81
- Braver TS, Barch DM, Gray JR, Molfese DL, Snyder A. 2001. Anterior cingulate cortex and response conflict: effects of frequency, inhibition and errors. *Cereb. Cortex* 11:825–36
- Bretherton I, Beeghly M. 1982. Talking about internal states: the acquisition of an explicit theory of mind. *Dev. Psychol.* 18:906–21
- Bretherton I, Fritz J, Zahn-Waxler C, Ridgeway D. 1986. Learning to talk about emotion: a functionalist perspective. *Child Dev.* 57:529–48
- Brooks R, Meltzoff AN. 2002. The importance of eyes: how infants interpret adult looking behavior. *Dev. Psychol.* 38:958–66
- Brunet E, Sarfati Y, Hardy-Bayle MC, Decety J. 2000. A PET investigation of the attribution of intentions with a nonverbal task. *NeuroImage* 11:157–66
- Buccino G, Binkofski F, Fink GR, Fadiga L, Fogassi L, et al. 2001. Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *Eur. J. Neurosci.* 13:400–4
- Burgund ED, Kang HC, Kelly JE, Buckner RL, Snyder AZ, et al. 2002. The feasibility of a common stereotactic space for children and adults in fMRI studies of development. *NeuroImage* 17:184–200
- Calder AJ, Lawrence AD, Keane J, Scott SK, Owen AM, et al. 2002. Reading the mind from eye gaze. *Neuropsychologia* 40:1129–38
- Canli T, Amin Z. 2002. Neuroimaging of emotion and personality: scientific evidence and ethical considerations. *Brain Cogn.* 50:414–31
- Caplan D. 2001. Functional neuroimaging studies of syntactic processing. *J. Psycholinguist. Res.* 30:297–320
- Cardinal RN, Parkinson JA, Hall J, Everitt BJ. 2002. Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. *Neurosci. Biobehav. Rev.* 26:321–52
- Carlson SM, Moses LJ. 2001. Individual differences in inhibitory control and children's theory of mind. *Child Dev.* 72:1032–53
- Caron AJ, Butler S, Brooks R. 2002. Gaze following at 12 and 14 months: Do the eyes matter? *Br. J. Dev. Psychol.* 20:225–39
- Caron AJ, Caron RF, MacLean DJ. 1988. Infant discrimination of naturalistic emotional expressions: the role of face and voice. *Child Dev.* 59:604–16
- Carpenter M, Akhtar N, Tomasello M. 1998a. Fourteen through 18-month-old infants differentially imitate intentional and accidental actions. *Infant Behav. Dev.* 21:315–30
- Carpenter M, Nagell K, Tomasello M. 1998b. Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monogr. Soc. Res. Child Dev.* 63:1–143
- Carruthers P, Smith PK, eds. 1996. *Theories of Theories of Mind*. New York/London: Cambridge Univ. Press. 390 pp.
- Casey BJ, Thomas KM, McCandliss B. 2001. Applications of magnetic resonance imaging to the study of development. In *Handbook of Developmental Cognitive Neuroscience*, ed. CA Nelson, M Luciana, pp. 137–48. Cambridge, MA: MIT Press
- Castelli F, Frith C, Happe F, Frith U. 2002. Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes. *Brain* 125:1839–49
- 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* 12:314–25
- Chaminade T, Meltzoff AN, Decety J. 2002. Does the end justify the means? A PET exploration of the mechanisms involved in human imitation. *NeuroImage* 15:318–28
- Corbetta M, Kincade JM, Shulman GL. 2002. Neural systems for visual orienting and their relationships to spatial working memory. *J. Cogn. Neurosci.* 14:508–23
- Corkum J, Moore C. 1995. Development of

- joint visual attention in infants. In *Joint Attention: Its Origin and Role In Development*, ed. C Moore, PJ Dunham, pp. 61–85. Hillsdale, NJ: Erlbaum
- Csibra G. 2003. Teleological and referential understanding of action in infancy. *Philos. Trans. R. Soc. London Ser. B* 358:447–58
- Cutting AL, Dunn J. 1999. Theory of mind, emotion understanding, language and family background: individual differences and interrelations. *Child Dev.* 57:529–48
- Decety J, Chaminade T. 2003. Neural correlates of feeling sympathy. *Neuropsychologia* 41:127–38
- Decety J, Chaminade T, Grezes J, Meltzoff AN. 2002. A PET exploration of the neural mechanisms involved in reciprocal imitation. *NeuroImage* 15:265–72
- Decety J, Grezes J. 1999. Neural mechanisms subserving the perception of human actions. *Trends Cogn. Sci.* 3:172–78
- Dehaene-Lambertz G, Dehaene S, Hertz-Pannier L. 2002. Functional neuroimaging of speech perception in infants. *Science* 298:2013–15
- Dennet D. 1996. *Kinds of Minds: Towards an Understanding of Consciousness*. New York/San Francisco: HarperCollins. 184 pp.
- Dornick D. 1978. Beliefs about beliefs. *Behav. Brain Sci.* 1:568–70
- de Villiers J. 2000. Language and theory of mind: What are the developmental relationships? See Baron-Cohen et al. 2000, pp. 83–123
- de Villiers J, de Villiers PA. 2000a. Linguistic determinism and false belief. In *Children's Reasoning and the Mind*, ed. P Mitchell, K Riggs. Hove, Engl: Psychology Press
- de Villiers J, de Villiers PA. 2000b. Linguistic determinism and the understanding of false beliefs. See Mitchell & Riggs 2000, pp. 191–228
- de Villiers J, Pyers JE. 2002. Complements to cognition: a longitudinal study of the relationship between complex syntax and false-belief-understanding. *Cogn. Dev.* 17:1037–60
- Doherty MJ, Anderson JR. 1999. A new look at gaze: preschool children's understanding of eye direction. *Cogn. Dev.* 14:549–71
- Downar J, Crawley AP, Mikulis DJ, Davis KD. 2002. A cortical network sensitive to stimulus salience in a neutral behavioral context across multiple sensory modalities. *J. Neurophysiol.* 87:615–20
- Dunn J, Brown J. 2001. Emotion, pragmatics and developments in emotion understanding in the preschool year. In *Jerome Bruner: Language, Culture, Self*, ed. D Bakhurst, S Shanker, pp. 88–103. Thousand Oaks, CA: Sage
- Eriksen BA, Eriksen CW. 1974. Effects of noise letters upon the identification of a target letter in a nonsearch task. *Percept. Psychophys.* 16:143–49
- Farrer C, Franck N, Georgieff N, Frith CD, Decety J, Jeannerod M. 2003. Modulating the experience of agency: a positron emission tomography study. *NeuroImage* 18:324–33
- Farrer C, Frith CD. 2002. Experiencing oneself vs another person as being the cause of an action: the neural correlates of the experience of agency. *NeuroImage* 15:596–603
- Feinman S, ed. 1992. *Social Referencing and the Social Construction of Reality in Infancy*. New York: Plenum. 424 pp.
- Ferstl EC, von Cramon DY. 2002. What does the frontomedian cortex contribute to language processing: coherence or theory of mind? *NeuroImage* 17:1599–612
- Field TM, Walden TA. 1982. Production and perception of facial expressions in infancy and early childhood. *Adv. Child Dev. Behav.* 16:169–211
- Flavell JH. 1988. The development of children's knowledge about the mind: from cognitive connections to mental representations. In *Developing Theories of Mind*, eds. JW Astington, PL Harris, DR Olson, pp. 244–67. New York/London: Cambridge Univ. Press
- Flavell JH. 1999. Cognitive development: children's knowledge about the mind. *Annu. Rev. Psychol.* 50:21–45
- Fletcher PC, Happe F, Frith U, Baker SC, Dolan RJ, et al. 1995. Other minds in the brain: a functional imaging study of "theory of mind"

- in story comprehension. *Cognition* 57:109–28
- Friederici AD. 2001. Syntactic, prosodic, and semantic processes in the brain: evidence from event-related neuroimaging. *J. Psycholinguist. Res.* 30: 237–50
- Frith CD, Frith U. 2000. The physiological basis of theory of mind: functional neuroimaging studies. See Baron-Cohen et al. 2000, pp. 335–56
- Frith U. 2001. Mind blindness and the brain in autism. *Neuron* 32:969–79
- Frith U, Frith CD. 2003. Development and neurophysiology of mentalizing. *Philos. Trans. R. Soc. London Ser. B* 358:459–73
- Gallagher HL, Frith CD. 2003. Functional imaging of “theory of mind.” *Trends Cogn. Sci.* 7:77–83
- Gallagher HL, Happe F, Brunswick N, Fletcher PC, Frith U, Frith CD. 2000. Reading the mind in cartoons and stories: an fMRI study of “theory of mind” in verbal and nonverbal tasks. *Neuropsychologia* 38:11–21
- Gallagher HL, Jack AI, Roepstorff A, Frith CD. 2002. Imaging the intentional stance in a competitive game. *NeuroImage* 16:814–21
- Gallese V. 2003. The manifold nature of interpersonal relations: the quest for a common mechanism. *Philos. Trans. R. Soc. London Ser. B* 358:517–28
- Gallese V, Goldman A. 1998. Mirror neurons and the simulation theory of mind-reading. *Trends Cogn. Sci.* 2:493–501
- Garavan H, Ross TJ, Murphy K, Roche RA, Stein EA. 2002. Dissociable executive functions in the dynamic control of behavior: inhibition, error detection, and correction. *NeuroImage* 17:1820–29
- Geday J, Gjedde A, Boldsen AS, Kupers R. 2003. Emotional valence modulates activity in the posterior fusiform gyrus and inferior medial prefrontal cortex in social perception. *NeuroImage* 18:675–84
- Gergely G, Csibra G. 1997. Teleological reasoning in infancy: the infant’s naive theory of rational action. A reply to Premack and Premack. *Cognition* 63:227–33
- Goel V, Grafman J, Sadato N, Hallett M. 1995. Modeling other minds. *NeuroReport* 6:1741–46
- Gopnik A, Astington JW. 1988. Children’s understanding of representational change and its relation to the understanding of false belief and the appearance-reality distinction. *Child Dev.* 59:26–37
- Greene J, Haidt J. 2002. How (and where) does moral judgment work? *Trends Cogn. Sci.* 6:517–23
- Grezes J, Decety J. 2001. Functional anatomy of execution, mental simulation, observation, and verb generation of actions: a meta-analysis. *Hum. Brain Mapp.* 12:1–19
- Halgren E, Rajj T, Marinkovic K, Jousmaki V, Hari R. 2000. Cognitive response profile of the human fusiform face area as determined by MEG. *Cereb. Cortex* 10:69–81
- Hamann S. 2003. Nosing in on the emotional brain. *Nat. Neurosci.* 6:106–8
- Happe F. 1992. Communicative competence and theory of mind in autism: a test of relevance theory. *Cognition* 48:101–19
- Harris PL. 1989. *Children and Emotion: The Development of Psychological Understanding*. Cambridge, MA: Basil Blackwell. 243 pp.
- Haxby JV, Hoffman EA, Gobbini MI. 2000. The distributed human neural system for face perception. *Trends Cogn. Sci.* 4:223–33
- Haxby JV, Hoffman EA, Gobbini MI. 2002. Human neural systems for face recognition and social communication. *Biol. Psychiatry* 51:59–67
- Heiser M, Iacoboni M, Maeda F, Marcus J, Mazziotta JC. 2003. The essential role of Broca’s area in imitation. *Eur. J. Neurosci.* 17:1123–28
- Hoffman EA, Haxby JV. 2000. Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nat. Neurosci.* 3:80–84
- Hood BM, Willen JD, Driver J. 1998. Adult’s eyes trigger shifts of visual attention in human infants. *Psychol. Sci.* 9:131–34
- Jellema T, Baker CI, Wicker B, Perrett DI.

2000. Neural representation for the perception of the intentionality of actions. *Brain Cogn.* 44:280–302
- Jiang Y, Kanwisher N. 2003. Common neural substrates for response selection across modalities and mapping paradigms. *J. Cogn. Neurosci.* In press
- Johnson SC. 2000. The recognition of mentalistic agents in infancy. *Trends Cogn. Sci.* 4:22–28
- Johnson SC. 2003. Detecting agents. *Philos. Trans. R. Soc. London Ser. B* 358:517–28
- Johnson SC, Booth A, O’Hearn K. 2001. Inferring the goals of a non-human agent. *Cogn. Dev.* 16:637–56
- Johnson SC, Slaughter V, Carey S. 1998. Whose gaze will infants follow? Features that elicit gaze-following in 12-month-olds. *Dev. Sci.* 1:233–38
- Just MA, Carpenter PA, Keller TA, Eddy WF, Thulborn KR. 1996. Brain activation modulated by sentence comprehension. *Science* 274:114–16
- Kanwisher N, McDermott J, Chun MM. 1997. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J. Neurosci.* 17:4302–11
- Konishi S, Nakajima K, Uchida I, Kikyo H, Kameyama M, Miyashita Y. 1999. Common inhibitory mechanism in human inferior prefrontal cortex revealed by event-related functional MRI. *Brain* 122:981–91
- Koski L, Wohlschlagler A, Bekkering H, Woods RP, Dubeau MC, et al. 2002. Modulation of motor and premotor activity during imitation of target-directed actions. *Cereb. Cortex* 12:847–55
- Lagattuta KH, Wellman HM. 2001. Thinking about the past: early knowledge about links between prior experience, thinking, and emotion. *Child Dev.* 72:82–102
- Lalonde CE, Chandler MJ. 2002. Children’s understanding of interpretation. *New Ideas Psychol.* 20:163–98
- LeDoux JE. 2000. Emotion circuits in the brain. *Annu. Rev. Neurosci.* 23:155–84
- Legerstee M, Barillas Y. 2003. Sharing attention and pointing to objects at 12 months: Is the intentional stance implied? *Cogn. Dev.* 18:91–110
- Leslie A. 1994. A theory of ToMM, ToBy, and Agency: core architecture and domain specificity. In *Mapping the Mind: Domain Specificity in Cognition and Culture*, ed. L Hirschfeld, S Gelman, pp. 119–48. New York/London: Cambridge Univ. Press
- Leslie A. 2000. “Theory of Mind” as a mechanism of selective attention. In *The New Cognitive Neurosciences*, ed. M Gazzaniga, pp. 1235–47. Cambridge, MA: MIT Press
- Leslie A, Polizzi P. 1998. Inhibitory processing in the false belief task: two conjectures. *Dev. Sci.* 1:247–53
- Leslie A, Thaiss L. 1992. Domain specificity in conceptual development. *Cognition* 43:225–51
- Malle BF. 2001. Folk explanations of intentional action. See Malle et al. 2001, pp. 265–86
- Malle BF. 2003. The relation between language and theory of mind in development and evolution. In *The Evolution of Language from Pre-Language*, ed. T Givon, BF Malle. Amsterdam/Erdenheim, PA: Benjamins. In press
- Malle BF, Knobe J. 2001. The distinction between desire and intention: a folk-conceptual analysis. See Malle et al. 2001, pp. 45–67
- Malle BF, Moses LJ, Baldwin DA, eds. 2001. *Intentions and Intentionality: Foundations of Social Cognition*. Cambridge, MA: MIT Press
- McCabe K, Houser D, Ryan L, Smith V, Trouard T. 2001. A functional imaging study of cooperation in two-person reciprocal exchange. *Proc. Natl. Acad. Sci. USA* 98:11832–35
- McCarthy G, Puce A, Gore JC, Allison T. 1997. Face-specific processing in the human fusiform gyrus. *J. Cogn. Neurosci.* 9:605–10
- Meltzoff AN. 1995. Understanding the intentions of others: re-enactment of intended acts by 18-month-old children. *Dev. Psychol.* 31:838–50
- Meltzoff AN, Decety J. 2003. What imitation tells us about social cognition: a rapprochement between developmental psychology

- and cognitive neuroscience. *Philos. Trans. R. Soc. London Ser. B* 358:491–500
- Milham MP, Banich MT, Webb A, Barad V, Cohen NJ, et al. 2001. The relative involvement of anterior cingulate and prefrontal cortex in attentional control depends on nature of conflict. *Brain Res. Cogn. Brain Res.* 12: 467–73
- Mitchell P, Riggs KJ, eds. 2000. *Children's Reasoning and the Mind*. Hove, Engl.: Psychology Press/Taylor & Francis
- Moll J, de Oliveira-Souza R, Eslinger PJ, Bramati IE, Mourao-Miranda J, et al. 2002. The neural correlates of moral sensitivity: a functional magnetic resonance imaging investigation of basic and moral emotions. *J. Neurosci.* 22:2730–36
- Morris JS. 2002. How do you feel? *Trends Cogn. Sci.* 6:317–19
- Moses LJ. 2001. Executive accounts of theory-of-mind development. *Child Dev.* 72:688–90
- Moses LJ, Baldwin DA, Rosicky JG, Tidball G. 2001. Evidence for referential understanding in the emotions domain at twelve and eighteen months. *Child Dev.* 72:718–35
- Narumoto J, Okada T, Sadato N, Fukui K, Yonekura Y. 2001. Attention to emotion modulates fMRI activity in human right superior temporal sulcus. *Brain Res. Cogn. Brain Res.* 12:225–31
- Nelson CA. 1987. The recognition of facial expressions in the first two years of life: mechanisms of development. *Child Dev.* 58:889–909
- Nelson K. 1996. *Language in Cognitive Development: Emergence of the Mediated Mind*. New York/London: Cambridge Univ. Press. 432 pp.
- Nichols S, Stich S, Leslie A, Klein D. 1996. Varieties of off-line simulation. In *Theories of Theories of Mind*, eds. P Carruthers, PK Smith, pp. 39–74. New York/London: Cambridge Univ. Press
- Pellicano E, Rhodes G. 2003. The role of eye-gaze in understanding other minds. *Br. J. Dev. Psychol.* 21:33–43
- Pelphrey KA, Singerman JD, Allison T, McCarthy G. 2003. Brain activation evoked by perception of gaze shifts: the influence of context. *Neuropsychologia* 41:156–70
- Perner J. 1993. *Understanding the Representational Mind*. Cambridge, MA: MIT Press
- Perner J, Lang B. 2000. Theory of mind and executive function: Is there a developmental relationship? See Baron-Cohen et al. 2000, pp. 150–81
- Perner J, Stummer S, Lang B. 1999. Executive functions and theory of mind: cognitive complexity or functional dependence? In *Developing Theories of Intention: Social Understanding and Self-Control*, ed. PD Zelazo, JW Astington, pp. 133–52. Mahwah, NJ: Erlbaum
- Peterson CC, Siegal M. 1995. Deafness, conversation and theory of mind. *J. Child Psychol. Psychiatry* 36:459–74
- Phillips AT, Wellman HM, Spelke ES. 2002. Infants' ability to connect gaze and emotional expression to intentional action. *Cognition* 85:53–78
- Povinelli DJ. 2001. On the possibilities of detecting intentions prior to understanding them. See Malle et al. 2001, pp. 225–48
- Premack D, Woodruff G. 1978. Does the chimpanzee have a theory of mind? *Behav. Brain Sci.* 1:515–26
- Preston SD, de Waal FBM. 2002. Empathy: its ultimate and proximate bases. *Behav. Brain Sci.* 25:1–72
- Puce A, Allison T, Bentin S, Gore JC, McCarthy G. 1998. Temporal cortex activation in humans viewing eye and mouth movements. *J. Neurosci.* 18:2188–99
- Puce A, Perrett D. 2003. Electrophysiology and brain imaging of biological motion. *Philos. Trans. R. Soc. London Ser. B* 358:435–45
- Repacholi BM, Gopnik A. 1997. Early reasoning about desires: evidence from 14- and 18-month-olds. *Dev. Psychol.* 33:12–21
- Rieffe C, Terwogt MM, Koops W, Stegge H, Oomen A. 2001. Preschooler's appreciation of uncommon desires and subsequent emotions. *Br. J. Dev. Psychol.* 19:259–74
- Rizzolatti G, Fogassi L, Gallese V. 2001. Neurophysiological mechanisms underlying the

- understanding and imitation of action. *Nat. Rev. Neurosci.* 2:661–70
- Roder B, Stock O, Neville H, Bien S, Rosler F. 2002. Brain activation modulated by the comprehension of normal and pseudo-word sentences of different processing demands: a functional magnetic resonance imaging study. *NeuroImage* 15:1003–14
- Ruby P, Decety J. 2001. Effect of subjective perspective taking during simulation of action: a PET investigation of agency. *Nat. Neurosci.* 4:546–50
- Ruffman T, Slade L, Rowlandson K, Rumsey C, Garnham A. 2003. How language relates to belief, desire and emotion understanding. *Cogn. Dev.* 113:1–20
- Saxe R, Kanwisher N. 2003. People thinking about thinking people: fMRI investigations of theory of mind. *NeuroImage*. In press
- Scholl BJ, Leslie AM. 2001. Minds, modules, and meta-analysis. Commentary on “Meta-analysis of theory-of-mind development: the truth about false belief.” *Child Dev.* 72:696–701
- Schultz RT, Grelotti DJ, Klin A, Kleinman J, Van der Gaag C, et al. 2003. The role of the fusiform face area in social cognition: implications for the pathobiology of autism. *Philos. Trans. R. Soc. London Ser. B* 358:415–27
- Searle JR. 1983. *Intentionality: An Essay in the Philosophy of Mind*. New York/London: Cambridge Univ. Press
- Shulman GL, d’Avossa G, Tansy AP, Corbetta M. 2002. Two attentional processes in the parietal lobe. *Cereb. Cortex* 12:1124–31
- Siegal M, Varley R. 2002. Neural systems involved in “theory of mind.” *Nat. Rev. Neurosci.* 3:463–71
- Stich S, Nichols S. 1998. Theory theory to the max. *Mind Lang.* 13:421–49
- Stroop JR. 1938. Factors affecting speed in serial verbal reactions. *Psychol. Monogr.* 50:38–48
- Sylvester CY, Wager TD, Lacey SC, Hernandez L, Nichols TE, et al. 2003. Switching attention and resolving interference: fMRI measures of executive functions. *Neuropsychologia* 41:357–70
- Tager-Flusberg H, Sullivan K. 2000. A componential view of theory of mind: evidence from Williams syndrome. *Cognition* 76:59–90
- Terwogt MM, Stegge H. 1998. Children’s perspective on the emotion process. In *The Social Child*, ed. A Campbell, S Muncer, pp. 249–69. Hove, Engl.: Psychology Press/Erlbaum
- Tomasello M, Strosberg R, Akhtar N. 1996. Eighteen-month-old children learn words in nonostensive contexts. *J. Child Lang.* 23:157–76
- Tootell RB, Reppas JB, Kwong KK, Malach R, Born RT, et al. 1995. Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. *J. Neurosci.* 15:3215–30
- Tremoulet PD, Feldman J. 2000. Perception of animacy from the motion of a single dot. *Perception* 29:943–51
- Vandenberghe R, Nobre AC, Price CJ. 2002. The response of left temporal cortex to sentences. *J. Cogn. Neurosci.* 14:550–60
- Varley R, Siegal M. 2000. Evidence for cognition without grammar from causal reasoning and “theory of mind” in an agrammatic aphasic patient. *Curr. Biol.* 10:723–26
- Varley R, Siegal M, Want SC. 2001. Severe impairment in grammar does not preclude theory of mind. *Neurocase* 7:489–93
- Vogeley K, Bussfeld P, Newen A, Herrmann S, Happe F, et al. 2001. Mind reading: neural mechanisms of theory of mind and self-perspective. *NeuroImage* 14:170–81
- Vuilleumier P, Armony JL, Clarke K, Husain M, Driver J, Dolan RJ. 2002. Neural response to emotional faces with and without awareness: event-related fMRI in a parietal patient with visual extinction and spatial neglect. *Neuropsychologia* 40:2156–66
- Watson AC, Painter KM, Bornstein MH. 2001. Longitudinal relations between 2-year-olds’ language and 4-year-olds’ theory of mind. *J. Cogn. Dev.* 2:449–57
- Wellman HM, Cross D. 2001. Theory of mind and conceptual change. *Child Dev.* 72:702–7
- Wellman HM, Cross D, Watson J. 2001.

- Meta-analysis of theory-of-mind development: the truth about false belief. *Child Dev.* 72:655–84
- Wellman HM, Harris PL, Banerjee M, Sinclair A. 1995. Early understanding of emotion: evidence from natural language. *Cogn. Emot.* 9:117–49
- Wellman HM, Lagattuta KH. 2000. Developing understandings of mind. See Baron-Cohen et al. 2000, pp. 21–49
- Wellman HM, Woolley JD. 1990. From simple desires to ordinary beliefs: the early development of everyday psychology. *Cognition* 35:245–75
- Whalen PJ. 1999. Fear, vigilance and ambiguity: initial neuroimaging studies of the human amygdala. *Curr. Dir. Psychol. Sci.* 7:177–87
- Wicker B, Michel F, Henaff MA, Decety J. 1998. Brain regions involved in the perception of gaze: a PET study. *NeuroImage* 8:221–27
- Wicker B, Perrett DI, Baron-Cohen S, Decety J. 2003. Being the target of another's emotion: a PET study. *Neuropsychologia* 41:139–46
- Wimmer H, Perner J. 1983. Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition* 13:103–28
- Winston JS, Strange BA, O'Doherty J, Dolan RJ. 2002. Automatic and intentional brain responses during evaluation of trustworthiness of faces. *Nat. Neurosci.* 5:277–83
- Wolpert DM, Doya K, Kawato M. 2003. A unifying computational framework for motor control and social interaction. *Philos. Trans. R. Soc. London Ser. B* 358:593–602
- Woodward AL. 1998. Infants selectively encode the goal object of an actor's reach. *Cognition* 69:1–34
- Woolfe T, Want SC, Siegal M. 2002. Signposts to development: theory of mind in deaf children. *Child Dev.* 73:718–78
- Zacks JM, Braver TS, Sheridan MA, Donaldson DI, Snyder AZ, et al. 2001. Human brain activity time-locked to perceptual event boundaries. *Nat. Neurosci.* 4:651–55
- Zaitchik D. 1990. When representations conflict with reality: the preschooler's problem with false beliefs and "false" photographs. *Cognition* 35:41–68
- Zaitchik D. 1991. Is only seeing really believing? Sources of the true belief in the false belief task. *Cogn. Dev.* 6:91–103