◆ Human Brain Mapping 30:2313–2335 (2009) ◆

REVIEW ARTICLE

Are There Theory of Mind Regions in the Brain? A Review of the Neuroimaging Literature

Sarah J. Carrington[1,2]* and Anthony J. Bailey[1]

[1]*Department of Psychiatry, The University of Oxford, Warneford Hospital, Headington, Oxford, United Kingdom*
[2]*The University of Oxford, FMRIB Centre, John Radcliffe Hospital, Headington, Oxford, United Kingdom*

Abstract: There have been many functional imaging studies of the brain basis of theory of mind (ToM) skills, but the findings are heterogeneous and implicate anatomical regions as far apart as orbitofrontal cortex and the inferior parietal lobe. The functional imaging studies are reviewed to determine whether the diverse findings are due to methodological factors. The studies are considered according to the paradigm employed (e.g., stories vs. cartoons and explicit vs. implicit ToM instructions), the mental state(s) investigated, and the language demands of the tasks. Methodological variability does not seem to account for the variation in findings, although this conclusion may partly reflect the relatively small number of studies. Alternatively, several distinct brain regions may be activated during ToM reasoning, forming an integrated functional "network." The imaging findings suggest that there are several "core" regions in the network—including parts of the prefrontal cortex and superior temporal sulcus—while several more "peripheral" regions may contribute to ToM reasoning in a manner contingent on relatively minor aspects of the ToM task. *Hum Brain Mapp 30:2313–2335, 2009.*

Key words: social cognition; mental states; fMRI; neural network; medial prefrontal cortex

INTRODUCTION

Theory of Mind (ToM)—the ability to think about mental states, such as thoughts and beliefs, in oneself and others [Premack and Woodruff, 1978]—underlies social interaction and allows people to make sense of the behavior of others. ToM is a complex cognitive function that requires integration of information from many sources. Two theories attempt to explain the psychological processes underlying ToM. The Theory Theory (TT) postulates that a set of causal laws relating external states, internal states, and behaviors are used to construct theories about the mental states of others [Gallese and Goldman, 1998]. The Simulation Theory (ST) suggests that the mental states of others are simulated using the same mental mechanisms involved in experiencing each state oneself [Gallese and Goldman, 1998; Ramnani and Miall, 2004; Williams et al., 2001]. Furthermore, it has been proposed that the simulation of mental states [Gallese and Goldman, 1998] may be supported by mirror neurons, which were first identified in non-human primates [di Pellegrino et al., 1992; Rizzolatti et al., 1996]. The theory and simulation theories of

*Correspondence to: Sarah Carrington, Department of Psychiatry, The University of Oxford, Warneford Hospital, Headington, Oxford, OX3 7JX, United Kingdom.
E-mail: sarah.carrington@sjc.ox.ac.uk

Received for publication 30 August 2007; Revised 24 July 2008; Accepted 27 August 2008
DOI: 10.1002/hbm.20671
Published online 25 November 2008 in Wiley InterScience (www.interscience.wiley com).

ToM need not be mutually exclusive: it is plausible that the more cognitively demanding theory theory may be adopted when simulation is inappropriate.

As with many cognitive functions, it is likely that ToM may also have a localized neurobiological basis. On the basis of data from single neuron recordings in non-human primates, Brothers [1990] argued that the orbitofrontal cortex (OFC), the superior temporal sulcus (STS), and the amygdala were dedicated—although not exclusively so—to primate social cognition, forming a "social brain." Brothers also suggested that the role of inferotemporal cortical regions, including the temporal pole, and the cingulate gyrus should be investigated. Brothers defined social cognition as "the processing of any information which culminates in the accurate perception of the dispositions and intentions of other individuals." This definition is more straightforward and all-encompassing than the traditional definition of ToM; moreover, it is limited to the perception of "dispositions and intentions," as only these perceptions are common to both human and non-human primates. Furthermore, Brothers explicitly incorporated the detection of eye gaze and affective facial expressions in her definition. In this review, social cognition is defined as the ability to understand people's behavior through the use of cues such as facial expression, eye gaze, body postures—including gesture—and social linguistic factors, such as prosody and the social content of speech. Thus, the current definition of ToM is encompassed within the term social cognition, but ToM is considered distinct in that it refers explicitly to individuals' mental states. Although eye gaze and facial expression may be used to guide interpersonal interactions, they do not necessarily involve the consideration of mental states.

The proposal that ToM has a neurobiological basis is supported by evidence of impaired ToM in individuals with autism [e.g. Baron-Cohen et al., 1985, 1986, 1999; Perner et al., 1989]. Although autism is a neurodevelopmental disorder, there is no neurobiological feature that is both universal and unique to autism [e.g. Bailey et al., 1998]; consequently, neuroanatomical examination of the brains of individuals with autism has so far not contributed significantly to our understanding of the neurobiological basis of impaired ToM. Acquired deficits in ToM following brain injury in adulthood have, however, indicated some regions of the brain that may be involved in ToM processing. For example, both frontal [Bach et al., 1998; Channon and Crawford, 2000; Happe et al., 2001; Rowe et al., 2001; Shamay-Tsoory et al., 2005; Stone et al., 1998; Stuss et al., 2001] and amygdalar damage [Shaw et al., 2004] have been associated with impaired ToM processing.

Many imaging studies of typically developing adults, employing positron emission tomography (PET) and functional magnetic resonance imaging (fMRI), have attempted to identify the neurobiological basis of ToM, but the findings (Table I) are heterogeneous, implicating regions as anatomically distant as orbitofrontal cortices [e.g. Baron-Cohen et al., 1994; Kozel et al., 2004] and the inferior parietal lobe [e.g. Fletcher et al., 1995; Gallagher et al., 2000]. Nevertheless, a pattern emerges when activated anatomical regions are grouped according to proximity (Table II). The medial prefrontal (mPFC) and orbitofrontal (OFC) region was implicated in nearly all (93%) studies, leading some authors to conclude that this region is "critical" for ToM [e.g. Gallagher et al., 2000]. Consistent with Brothers' notion of a "social brain," the anterior temporal lobe,—encompassing the amygdala—and superior temporal regions were associated with ToM reasoning in 38% and 50% of studies, respectively. In addition, the anterior- and paracingulate cortices were activated in 55% of studies and the temporoparietal junction in 58%.

The diverse methodology employed in the ToM imaging studies raises the question of whether variation in the regions activated by ToM reasoning is due to task variables, such as the paradigm, the mental state(s) investigated, and task language demands. We consider each of these possibilities.

EXPERIMENTAL PARADIGMS

Functional imaging studies[1] of ToM have used many experimental paradigms, including the recognition of mental state terms [e.g. Baron-Cohen et al., 1994], stories [e.g. Fletcher et al., 1995; Vogeley et al., 2001], single-frame cartoons [e.g. Gallagher et al., 2000], comic strip cartoons [e.g. Brunet et al., 2000; Walter et al., 2004], and interactive games such as stone-paper-scissors [e.g. Gallagher et al., 2002].

Mental State Terms

The recognition of mental state terms, such as want, think, and believe was one of the first paradigms used in imaging studies. In a single photon emission computerized tomography (SPECT) study, Baron-Cohen et al. [1994] asked participants to listen to two lists of words and decide whether each heard word was consistent with the theme of the list. One list included mostly mental state terms, whilst the control list contained words referring mainly to the body. The mental state terms elicited increased activity in right OFC and decreased activity in the left frontopolar region compared to the control words. Nevertheless, the region of interest (ROI) approach to data analysis included only anterior regions. In an analysis of whole-brain activity evoked whilst choosing mental state terms to describe the feelings conveyed by photographs of the eyes, Baron-Cohen et al. [1999] found activity in left frontal regions, including the dorsolateral (dl) PFC, medial frontal cortex (MFC), and supplementary motor area (SMA). Activity was also reported in bilateral temporoparietal regions, the insula, and left amygdalar, and hippocampal regions.

[1]Unless otherwise stated, studies employed fMRI.

TABLE I. A summary of the results from functional imaging studies investigating the neural correlates of ToM

	Regions																		
	MPFC (8/9/10)	LPFC (44/45/46/47)	SMA (6)	OFC (11)	Motor cortex (BA4)	ACC (24/32)	Paracingulate (10/32)	Precuneus (7/31)	PCC (23/31)	Temporal poles (38)	STS/STG (21/22)	MTG (21)	TPJ (39/40)	IPL (40)	Amygdala	Occipital cortex	Insula	Fusiform (36/37)	Cerebellum
Baron-Cohen et al., 1994	X			X															
Fletcher et al., 1995	X					X			X					X					
Goel et al., 1995	X									X		X							
Baron-Cohen et al., 1999	X	X	X			X					X	X	X		X		X		
Brunet et al., 2000	X	X	X			X	X			X		X				X			X
Castelli et al., 2000	X									X	X		X			X		X	
Gallagher et al., 2000 (S)	X									X			X					X	
Gallagher et al., 2000 (C)	X		X					X					X						
Sabbagh and Taylor, 2000	X																		
McCabe et al., 2001	X						X												
Spence et al., 2001	X	X	X			X								X					
Vogeley et al., 2001	X	X	X			X				X			X						
Gallagher et al., 2002	X	X					X												X
Lee et al., 2002	X	X	X								X			X					
Mitchell et al., 2002	X				X						X	X				X		X	
Calarge et al., 2003	X	X				X	X		X	X				X					X
Ganis et al., 2003	X							X										X	X
Saxe and Kanwisher, 2003	X							X			X		X						
German et al., 2004	X	X				X					X	X			X			X	X
Grèzes et al., 2004a		X		X		X					X				X			X	
Grèzes et al., 2004b	X			X			X				X	X					X		X
Kozel et al., 2004				X		X													X
Mason et al., 2004	X		X			X													
Rilling et al., 2004	X						X	X	X		X								
Walter et al., 2004	X						X												X
Iacoboni et al., 2005		X									X								
Mitchell et al., 2005a	X							X						X		X			
Mitchell et al., 2005b	X	X				X	X				X		X		X	X	X		
Mosconi et al., 2005											X	X	X						
Saxe and Powell, 2005	X								X		X		X		X				
Völlm et al., 2005	X			X					X	X	X	X	X			X		X	X
Mitchell et al., 2006	X	X				X										X			
Ciaramidaro et al., 2007							X	X					X						
Gobbini et al., 2007 (S)	X						X	X		X			X						
Gobbini et al., 2007 (A)	X	X				X		X		X	X		X				X	X	X
Kobayashi et al., 2007	X												X	X		X			
Sommer et al., 2007	X		X			X		X			X		X						
Mitchell, 2008	X								X				X						
Lissek et al., 2008	X	X				X		X			X		X				X		
Young and Saxe, 2008	X							X		X	X		X						
Total (40)	35	14	8	5	1	15	10	11	6	10	18	8	18	6	5	8	5	8	10
Percent	88	35	20	13	3	38	25	28	15	25	45	20	45	15	13	20	13	20	25

The regions in which ToM-related activity was observed are reported for each study in chronological order. The total number of studies implicating each region was calculated and reported as a percentage (MPFC, medial prefrontal cortex; LPFC, lateral prefrontal cortex; SMA, supplementary motor area; OFC, orbitofrontal cortex; ACC, anterior cingulate cortex; STS/STG, superior temporal sulcus/superior temporal gyrus; MTG, middle temporal gyrus; TPJ, temporoparietal gyrus; IPL, inferior parietal lobe). The results for the two different paradigms employed by Gallagher et al. (2000) and Gobbini et al. (2007) have been considered independently both here and in Table II. Although Kobayashi et al. also included two paradigms, they did not report the results for the two tasks separately; consequently, the ToM-related activity reported here is the activity that they reported common to both paradigms.

TABLE II. A summary of the results from functional imaging studies investigating the neural correlates of ToM grouped according to anatomical proximity

	Medial PFC and OFC	Lateral PFC	SMA and motor cortex	ACC and para-cingulate	Precuneus and PCC	Anterior temporal lobe	STS and surrounding cortex (STG/MTG)	Temporo-parietal junction (includes IPL)	Occipital cortex	Insula	Fusiform	Cerebellum
Baron-Cohen et al., 1994	X											
Fletcher et al., 1995	X			X	X			X				
Goel et al., 1995	X					X	X					
Baron-Cohen et al., 1999	X	X	X	X		X	X	X		X		
Brunet et al., 2000	X	X	X	X		X	X		X			X
Castelli et al., 2000	X					X	X	X	X		X	
Gallagher et al., 2000 (S)	X					X		X				
Gallagher et al., 2000 (C)	X		X		X			X			X	
Sabbagh and Taylor, 2000	X											
McCabe et al., 2001	X			X								
Spence et al., 2001	X	X	X	X				X				
Vogeley et al., 2001	X	X	X	X		X		X				
Gallagher et al., 2002	X	X		X								X
Lee et al., 2002	X	X	X				X	X				
Mitchell et al., 2002	X		X				X		X		X	
Calarge et al., 2003	X	X		X	X	X		X				X
Ganis et al., 2003	X				X						X	X
Saxe and Kanwisher, 2003	X				X		X	X				
German et al., 2004	X	X		X		X	X				X	X
Grèzes et al., 2004a	X	X		X		X	X				X	
Grèzes et al., 2004b	X			X			X			X		X
Kozel et al., 2004	X			X								X
Mason et al., 2004	X		X	X								
Rilling et al., 2004	X			X	X		X					
Walter et al., 2004	X			X								X
Iacoboni et al., 2005		X					X					
Mitchell et al., 2005a	X				X			X	X			
Mitchell et al., 2005b	X	X		X		X	X	X	X	X		
Mosconi et al., 2005							X	X				
Saxe and Powell, 2005	X				X	X	X	X				
Völlm et al., 2005	X				X	X	X	X	X		X	X
Mitchell et al., 2006	X	X		X					X			
Ciaramidaro et al., 2007				X	X			X				
Gobbini et al., 2007 (S)	X			X	X	X		X				
Gobbini et al., 2007 (A)	X	X		X	X	X	X	X		X	X	X
Kobayashi et al., 2007	X							X	X			
Mitchell, 2008	X				X			X				
Sommer et al., 2007	X		X	X	X		X	X				
Lissek et al., 2008	X	X		X	X		X	X		X		
Young and Saxe, 2008	X				X	X	X	X				
Total (40)	37	14	9	22	16	15	20	23	8	5	8	10
Percent	93	35	23	55	40	38	50	58	20	13	20	25

The brain regions listed in Table I are grouped here according to anatomical proximity to provide a more coherent picture of the brain basis of ToM. The proportion of studies implicating each particular region expressed as a percentage (PFC, prefrontal cortex; OFC, orbitofrontal cortex; SMA, supplementary motor area; ACC, anterior cingulate cortex; PCC, posterior cingulate cortex; STS, superior temporal cortex; STG/MTG, superior temporal gyrus/middle temporal gyrus; IPL, inferior parietal lobe). Note the absence of activity in the MPFC/OFC region defined here in the Ciaramidaro et al. (2007) study. The anterior paracingulate is within the medial prefrontal cortex; the distinction is made here to allow visualization of possible subdivisions within cortex.

Mason et al. [2004] also adopted a word-recognition approach but paired a target word ("human" or "dog") with a subsequent action word and asked participants whether the action word could be used to describe the target. Some actions were specific to either humans (e.g. talk) or dogs (e.g. bark), whilst others were not species-specific (e.g. walk and run). Mason et al. hypothesized that only actions associated with humans would automatically evoke attribution of mental states Indeed, Mason et al. found that the action words associated with humans evoked significant activations in the middle and medial frontal gyri, and in the right anterior cingulate cortex (ACC), compared with the words associated with dogs. Mitchell et al. [2005a] used similar target-adjective pairings, but rather than assuming that action words would elicit mental state attributions, the authors presented "psychological-state" words characteristic of both humans and dogs (e.g. "curious," "energetic"). Additional control conditions included the presentation of "body-part" words applicable to both humans and dogs, "abstract" words that were not particularly good descriptors for either target, and "object-part" words. Psychological-state judgments for both species elicited greater activity in the right dorsomedial (dm) PFC than judgments about body parts, indicating that activation of this region during mental state attribution is not confined to conspecifics.[2] Although the apparent species-specific activity evoked by action words [Mason et al., 2004] compared with the lack of specificity associated with mental state words [Mitchell et al., 2005a] appears counter-intuitive, the general tendency to anthropomorphize animals is probably associated with the attribution of "human" mental states to other species. Furthermore, Mason et al. argued that only human action descriptions would evoke the attribution of mental states, which would explain the species-specific patterns of activity. Despite the relative simplicity and similarity of these four paradigms, the only brain regions consistently associated with ToM were the medial prefrontal and orbitofrontal regions.

Simple Questions

The presentation of simple questions has also been used to investigate ToM, particularly, with respect to active deception, which requires consideration of others' knowledge and beliefs to mislead them. Comparison of the activity evoked by participants answering the same questions both truthfully and falsely, implicated several regions including the medial/orbital [Ganis et al., 2003; Kozel et al., 2004; Lee et al., 2002; Spence et al., 2001] and lateral [Lee et al., 2002; Spence et al., 2001] prefrontal regions, the cingulate cortex [Kozel et al., 2004; Lee et al., 2002; Spence et al., 2001], and the fusiform gyrus [Ganis et al., 2003], but only the mPFC/OFC region was activated in all four studies.

Stories

A paradigm commonly used to investigate ToM is based on the three categories of prose passages devised by Happé [1994] and adapted by Fletcher et al. [1995]: ToM, physical causality, and unlinked sentences. The only difference between the ToM and physical story conditions is the need to attribute mental states. Passages of unlinked sentences control for the integration of information for story comprehension. In the critical comparison of the ToM and physical story conditions, Fletcher et al. reported ToM-related activity in the left MFG, the ACC, posterior cingulate cortex (PCC), and the right IPL with only the left MFG responding exclusively during the ToM condition. Similarly, using adapted versions of Happé's stories, Gallagher et al. [2000] and Gobbini et al. [2007] reported ToM-related activity in the mPFC, as well as in the temporal poles and temporoparietal cortex. Gobbini et al. reported additional activity in both anterior and posterior regions of the cingulate gyrus bilaterally. Vogeley et al. [2001] observed a slightly more posterior focus of ToM-related activity, in the ACC, although activation did extend anteriorly into the mPFC. Vogeley et al. presented Happé's original stories and added a condition that required participants to imagine themselves in the context of the story. The attribution of mental states to oneself is inherent in Premack and Woodruff's definition of ToM but is investigated less frequently than the attribution of mental states to others. As well as activating the ACC, the "self" condition also evoked activity in the right TPJ and medial regions of the superior parietal lobe (SPL). The authors concluded that while the ACC was the agent-independent "cerebral implementation" of ToM capacity, additional, differentiable brain regions were associated with the attribution of mental states to oneself.

Saxe and Kanwisher [2003] devised four novel categories of stories, false belief, human action, non-human inferences, and mechanical inferences. Only the first two conditions required ToM reasoning and these conditions elicited greater activity in the anterior STS, precuneus, and bilaterally in the TPJ; the anterior STS and the TPJ were also activated in a condition assessing the participant's understanding of the protagonist's desire. In a second experiment with full-brain coverage, the authors demonstrated that compared with a false photograph control story,[3] false belief stories elicited activity in the right medial superior frontal gyrus and in the frontal pole [Saxe and Kanwisher, 2003]. The latter findings extended an earlier event-related potential (ERP) electrophysiology study, which implicated the left frontal lobe in a comparison of false belief and false photograph stories [Sabbagh and Taylor, 2000]. Consistent with Saxe and Kanwisher [2003], Saxe and Powell

[2]An earlier study of the neural substrates associated with person and object knowledge that used similar target-adjective pairings, [Mitchell et al., 2002] demonstrated that the brain regions associated with ToM in the studies discussed above were distinct from the regions associated with object knowledge.

[3]Based on the task devised by Zaitchik [1990].

[2006], Mitchell [2007] and Young and Saxe [2008] all reported significantly greater activity in several brain regions, including the TPJ bilaterally and regions bilaterally within the mPFC, for false belief stories compared with false photograph stories. Saxe and Powell [2006] then devised three new categories of stories to probe these regions of interest: (1) Appearance stories described the protagonist's physical and social appearance with no specific reference to internal states; (2) Bodily-sensation stories referred only to physical internal states and experiences, which the authors argued elicited only components of ToM thought to develop at a relatively young age, such as goals, perceptions, and feelings; (3) Thought stories briefly described a protagonist's beliefs and reasoning, which the authors argued required more sophisticated and arguably later-developing components of ToM. The thought stories evoked significantly increased activity bilaterally in the TPJ compared with the other two categories. The right supramarginal gyrus, cingulate cortex, and cerebellum were more active during the bodily-sensation stories compared with appearance stories, implicating these regions in the earlier-developing components of ToM. Nevertheless, the activity in the prefrontal regions identified by contrasting false belief and false photograph stories did not differentiate between the three conditions. Given that even the appearance stories had a social element, Saxe and Powell suggested that the mPFC may play a general role in the representation of socially relevant information about others, which might account for the consistent activation of the mPFC across the different studies.

Static Images

Several pictorial paradigms have been developed to investigate ToM. Gallagher et al. [2000] presented participants with both the stories described above, and single-frame cartoons belonging to the same three categories. Some of the cartoons required comprehension of the characters' mental states, whilst others did not, and some cartoons were simply "jumbled pictures," the pictorial equivalent of unlinked sentences. Compared with the control conditions, ToM cartoons elicited activity bilaterally in the mPFC, and in the right TPJ, precuneus and fusiform gyrus. An interaction analysis of condition (ToM vs. non-ToM) by task (stories or cartoons) revealed modality-independent ToM-related activity in bilateral mPFC only.

Kobayashi et al. [2007][4] conducted an investigation of modality-related ToM (stories versus cartoons) similar to that of Gallagher et al. [2000], although the two studies differed in several ways. Firstly, Kobayashi et al. presented five-frame cartoons depicting a sequence of events, rather than the single-frame cartoons used by Gallagher et al. Secondly, Kobayashi et al. presented both the stories and cartoons serially, either frame-by-frame or sentence-by-sentence, which was a mnemonically more demanding task than single-frame presentation. Thirdly, Kobayashi et al. focused on second-order false beliefs, whilst Gallagher at al. did not investigate one specific mental state. To represent second-order false beliefs pictorially, Kobayashi et al. illustrated the thought bubble of one person encompassing the thought bubble of a second person, adding complexity to the cartoon stimuli. Finally, only Kobayashi et al. directly compared the activity evoked by the ToM stories and cartoons. Interestingly, Kobayashi et al. failed to replicate Gallagher et al.'s finding that the mPFC was the only region uniquely associated with ToM reasoning; rather they observed modality-independent activity more dorsolaterally (dl) in the PFC, as well as in several more posterior regions, including the right IPL and bilateral TPJ. Furthermore, an ROI analysis of the Kobayashi data revealed that only activity in the bilateral TPJ and right IPL was specific for ToM compared with both the non-ToM and baseline conditions in both modalities. Kobayashi et al. suggested that the dlPFC activity in the ToM conditions may have reflected the additional inhibitory control demanded by the attribution of second-order false beliefs. These results, therefore, support previous suggestions that the TPJ may play a more central role in ToM than the mPFC [e.g. Saxe and Kanwisher, 2003; Saxe and Powell, 2006; Saxe and Wexler, 2005].

A sequence of cartoons illustrating true and false beliefs was also presented by Sommer et al. [2007]. The cartoon sequences followed the Sally-Ann format [Baron-Cohen et al., 1985] in which a change in the location of an object is made either with (true belief) or without (false belief) the critical protagonist's awareness. False belief cartoons evoked more activity than true belief cartoons in several regions, including the dorsal ACC, the PFC, and the right TPJ. The authors proposed non-ToM roles for both the ACC and PFC, suggesting that the TPJ was the only region specifically associated with ToM. Nevertheless, activation of the TPJ was not associated specifically with true belief or with both belief conditions combined. Sommer et al. suggested a more exclusive role for the TPJ when participants had to "decouple" their representation of the protagonist's beliefs from reality. Given that true beliefs do not contrast with reality, the lack of TPJ activity in this condition is understandable. Furthermore, Sommer et al. suggested that true belief scenarios could be resolved without mental state attribution, simply through the representation of reality. As such, it is perhaps unsurprising that the one region Sommer et al. implicated in ToM reasoning was not associated with the representation of true belief. Lissek et al. [2008] also presented participants with series of cartoons depicting interactions between characters. Participants were specifically asked to attribute both beliefs and intentions to the characters, and activity was contrasted

[4]The sample in this study included both adults and children, as one aim of the study was to determine age-related differences in the neural correlates of ToM. Developmental changes in the neural substrates of ToM are not considered in this review, and results reported in this article are those that were found to remain constant across the two age groups.

with conditions in which participants were asked about physical properties of the same series of cartoons that had been jumbled. Compared with the non-ToM condition, ToM sequences evoked activity in superior, inferior, and medial regions of the PFC, the ACC, TPJ, precuneus, and the insula.

A different pictorial paradigm eliciting ToM reasoning is the comic strip task [Brunet et al., 2000; Ciaramidaro et al., 2007; Vollm et al., 2006; Walter et al., 2004]. Three-frame cartoons depicting a short sequence of events are followed by three single frames illustrating alternative endings to the sequence; participants have to choose which single frame illustrates the most appropriate ending.[5] The comic strips fall into three conditions: one designed to evoke the attribution of intentions to the character, and two conditions depicting sequences of physical causality, one with and one without characters. Using these comic strip stimuli, Brunet et al. [2000] demonstrated that the attribution of intentions condition elicited activity in medial and inferior areas of the right PFC—including the ACC—anterior temporal regions bilaterally, and the left cerebellum when compared with both physical causality conditions. These findings were partially replicated in a study using the same cartoons for the intentions condition, an additional ToM condition,[6] but only some of the original physical causality cartoons [Vollm et al., 2006]. Compared with the physical causality condition, the ToM comic strips evoked increased activity in medial prefrontal and orbitofrontal regions, the TPJ, and the temporal cortex.

Walter et al. [2004] presented participants with comic strips involving more than one character and thus could distinguish between private intentions with one character—equivalent to the ToM conditions in the studies by Brunet et al. and Vollm et al.—private intentions with two characters, and communicative intentions. Although Walter et al. reported activity in "typical" ToM regions for strips involving private intentions, they concluded that the paracingulate cortex was selectively engaged when processing mental states, specifically intentions, associated with social interactions between several characters. Ciaramidaro et al. [2007] developed comic strip cartoons portraying private intentions; intentions with the potential to be shared and, therefore, social; and communicative intentions, which were both social and shared between two or more characters. In comparison with private intentions, both social intention conditions activated the anterior paracingulate. Thus, Ciaramidaro et al. extended the findings of Walter et al. by demonstrating that the paracingulate cortex was activated by intentions that were only potentially social. Both Walter et al. and Ciaramidaro et al. argued that their results were indicative of specialization within the neural system underlying ToM. Hence, several studies [Ciaramidaro et al., 2007; Sommer et al., 2007; Walter et al., 2004] suggest that subsets of ToM regions may subsume different aspects of ToM.

Goel et al. [1995] investigated whether a historical figure would have known the function of photographed objects presented to participants during PET imaging. Compared with more simple, nonmentalistic inferences, these ToM judgments evoked distributed activation that included prefrontal and temporal regions, although only the left orbitomedial frontal region was exclusively associated with ToM. Mitchell et al. [2005b] presented photographs of faces during fMRI and asked participants either how happy the person had been to be photographed (the ToM task), and/or how symmetrical the face was (the non-ToM task). Compared with judgments of symmetry, the ToM task was associated with increased activity bilaterally in the dmPFC and TPJ, in the right STS, and the left amygdala. Mitchell et al. extended these findings by demonstrating that the extent to which participants judged each face to be similar to their own was negatively correlated with activity in the dmPFC and positively correlated with activity in the vmPFC, implying a dissociation of ToM function within the mPFC. The authors suggested that vmPFC may support the attribution of mental states of similar others through the simulation of those mental states in oneself (Simulation Theory). By contrast, more dorsomedial prefrontal regions may support ToM reasoning when simulation is inappropriate, i.e. for dissimilar others (theory theory). Mitchell et al. [2006] further investigated ToM reasoning for similar and dissimilar others by pairing photographs of two "target" faces with a description of their political, religious, and social views. One target held views that were similar to those of the participant; the other had dissimilar views. During scanning, participants were asked to indicate on a four-point scale, the likelihood that the target presented would agree with opinion-related questions. Consistent with the suggested simulation role for the vmPFC [Mitchell et al., 2002], activity in this region was greater when participants made judgments for the similar target than for the dissimilar target. This pattern of activation was also seen in other frontal regions, including the cingulate, and bilaterally in the occipital cortex. Only the dmPFC, however, was more active during judgments about the dissimilar target, consistent with the idea of functional dissociation within the mPFC[7] [Mitchell et al., 2005b].

In summary, Mitchell et al., [2005b, 2006] argued that ventromedial prefrontal regions are recruited for simulation of the mental states of similar others, but that more dorsomedial regions are recruited when simulation is inap-

[5]Unlike the multi-frame cartoons used by Kobayashi et al. [2007] and Sommers et al. [2007], all three frames of these stimuli were presented simultaneously.

[6]This additional condition is discussed in more detail below.

[7]It should be noted, however, that the neural mechanism proposed to underlie the simulation of mental states has traditionally been associated with more lateral regions of the frontal cortex [e.g. Ehrsson et al., 2000; Iacoboni et al., 1999; Krams et al., 1998].

propriate and more complex mental state reasoning (such as the construction of theories—Theory Theory) must be implemented. Although it has been suggested that mirror neurons (MNs) mediate the simulation of mental states [Gallese and Goldman, 1998], the proposed ventral locus is inconsistent with reports of MN-like properties in more lateral regions of the frontal cortex [Ehrsson et al., 2000; Iacoboni et al., 1999; Krams et al., 1998]. The simulation proposal of Mitchell et al. in fact arises from previous reports of vmPFC involvement in self-referential thinking when reporting one's own "internal" states [Kelley et al., 2002; Macrae et al., 2004] or when adopting a first-person perspective [Vogeley et al., 2001]. Nevertheless, self-referential thinking should not be confused with simulation. It is possible that the differing activations associated with thinking about similar and dissimilar others may be due to a "like me" mental comparison, rather than simulation. Such a comparison would be more appropriate for similar than dissimilar others, and could, therefore account for the differences in activation reported by Mitchell et al. Furthermore, reports of ToM-related activity in the vmPFC in other studies using different paradigms [e.g. Gallagher et al., 2000, 2002; Vogeley et al., 2001] would not be inconsistent with this idea, as a "like me" comparison could potentially occur in any social situation. Thus, although the functional dissociation within the mPFC lends support to the idea of distinct subsets within a group of anatomical regions underlying ToM function, it does not appear that different ToM paradigms are associated with activation of distinct subsets of regions.

Animations and Videos

Some researchers have used animations to investigate ToM. Mosconi et al. [2005] showed 7- to 10-year-old children videos of animated characters who shifted their gaze either toward (congruent) or away from (incongruent) a flashing checkerboard presented in the periphery of the child's visual field. The authors predicted that the incongruent condition would evoke activity in regions of the brain associated with ToM, as this gaze shift should violate participants' expectations about the character's intentions. Increased activation was found in the posterior STS, the middle temporal gyrus (MTG), and the IPL of the right hemisphere for incongruent compared with congruent shifts of eye gaze. The Mosconi et al. study is one of only three reviewed studies that did not report activity in the mPFC/OFC region (Table II). Perceiving direction of eye gaze is central to the development of joint attention behaviors that are precursors of ToM [Baron-Cohen, 1995; Wicker et al., 1998]. Nevertheless, if the incongruent gaze condition involved a violation of expectation, then expectations about the character's intentions must also have been formed in the congruent conditions. Thus, although the increased mPFC/OFC activity in incongruent trials may be due to the greater processing demanded by violation of expectation, the lack of activity in the critical contrast may be because both conditions required mental state reasoning.

Castelli et al. [2000] developed a novel task, based on the silent animations of Heider and Simmel [1944], to encourage attribution of mental states to the kinematic properties of simple shapes. Participants observed animations of two triangles engaged in three types of interactions. The ToM condition involved interactions implying complex mental states, such as one triangle mocking or surprising the other. In goal-directed interactions, the purposeful actions of one shape determined the actions of the other. In the random motion condition, the shapes moved around the screen independently of each other and without interacting. Consistent with the studies reviewed above, the ToM animations evoked significantly greater activity in the mPFC, TPJ, temporal poles, and lateral superior occipital regions[8] than the control conditions. Using the same animations, Gobbini et al. [2007] contrasted activity evoked by the social animations with that evoked by the random motion animations. Consistent with Castelli et al. [2000], activity was seen in medial prefrontal regions, specifically the right anterior paracingulate cortex, and the temporal poles bilaterally. The contrast also revealed bilateral activity in the posterior STS and TPJ, as defined in this review. Activity did not extend into the more posterior portion of the TPJ reported by Castelli et al. [2000]. Gobbini et al. suggested that the animations could be understood in terms of the underlying goals of the action; that is, they distinguished between the intentions of an action and more abstract intentions, which they suggested would be represented by the pSTS and TPJ, respectively. Furthermore, Gobbini et al. proposed that the absence of activity in the TPJ indicated that the animations could be understood only in terms of the action goals, with relatively little reference to more abstract intentions.[9] The discrepancy between the patterns of activity reported by Castelli et al. and Gobbini et al. may simply be due to the definitions of the pSTS and TPJ employed by the two studies; given that the two studies employed the same task, it is unlikely that they tapped different mental states. The significance of the definition and size of brain regions is discussed below.

Videos of human actors have also been used to elicit mental state reasoning during functional neuroimaging.

[8]Castelli et al. suggested, however, that this task specific activation of superior lateral occipital regions may have reflected differences in local vs. global processing elicited by the different conditions rather than ToM *per se*.

[9]It should be noted, however, that the descriptions of the animations reported by Castelli et al. [2000] referred to abstract mental states such as "want" and "pretend." Consistent with these behavioral findings, Gobbini et al. [2007] reported activity in the anterior paracingulate region of the mPFC during observation of the social animations; because this region has been associated with ToM using other paradigms, it could be argued that the animations may have elicited more abstract mental state reasoning.

German et al. [2004] presented short videos of actors either performing or pretending to perform simple everyday actions such as reaching for a book. Half the videos were interrupted with a blue screen before completion of the action. Participants were asked to indicate whether each video was complete: judgments that were unrelated to mental states. German et al. hypothesized that observation of pretence would elicit more mental state attribution than viewing performance of the same actions, even when there was no instruction to attend to mental states. Consistent with this hypothesis, the observation of pretence evoked activity in several frontal regions previously implicated in ToM, including both medial and lateral PFC and the ACC, in addition to the posterior middle and superior temporal gyri, the fusiform gyrus, and the amygdala.

To investigate the brain basis of ToM, Iacoboni et al. [2005] presented three types of video clip: simple actions with no context or easily inferable intention, actions within context that facilitated the inference of intention (e.g. cleaning up or drinking), and clips displaying only the context with no action. Contrasting the intentions condition with simple actions revealed increased activity in the dorsal pars opercularis region of the right inferior frontal cortex that could not be attributed to the presence of objects in the videos. As the dorsal pars opercularis has been identified as exhibiting "mirror" properties [e.g. Ehrsson et al., 2000; Iacoboni et al., 1999; Koski et al., 2003; Krams et al., 1998], the authors proposed a role for mirror neurons in the attribution of intentions as well as in action recognition. The activity reported by Iacoboni et al. [2005] could be interpreted as supporting the simulation theory of ToM, but there is a distinction between the simulation of mental states, proposed by the simulation theory of ToM, and the simulation of motor actions. Although Iacoboni et al. suggested that mirror neurons may be involved in coding motor intentions—i.e. the intention or goal of an action—they did not suggest that mirror neurons code more abstract intentions.[10] Indeed, the study by Iacoboni et al. is the second reviewed study that did not report ToM-related activity in the mPFC, which would be understandable if the intentions represented in this study were simply motor intentions.

Grèzes et al. [2004b] presented videos of actors carrying boxes of known weight. In some videos, however, the actors had been misinformed about the weight of the boxes that they went to pick up, i.e. they had a false belief/expectation. Participants were asked to indicate whether they believed the actor had been correctly informed about the weight of the box i.e. whether they had a true or false belief. Compared with true belief, the attribution of false belief evoked significantly greater activity in inferior frontal regions, including the anterior paracingulate and dorsomedial regions, the STS, and the left cerebellum. Grèzes et al. did not report a contrast between the judgments of true belief and the null events used as a control, so no conclusions could be drawn regarding the neural underpinnings of true belief attribution. The authors argued that violation of expectations about the actors' movements in the false belief trials required participants to update their representations of the actors' mental states, thus inducing additional ToM related activity. Using similar video clips, Grèzes et al. [2004a] asked participants whether the actor was actively attempting to deceive the viewer about the weight of the box. Consistent with previously discussed studies of ToM [e.g. Baron-Cohen et al., 1999; German et al., 2004], activity in the ACC and amygdala was significantly increased when participants judged that they were being deceived. In both studies, it is possible that judgments were based on the nature of the actor's movements. Indeed, consistent with the functional dissociation between the TPJ and pSTS proposed by Gobbini et al., both studies reported activity in the pSTS rather than the TPJ. Nevertheless, participants were explicitly asked about the characters' mental states in each study. Furthermore, activation of regions such as the mPFC and ACC, which have both been associated with ToM using other paradigms, suggests that the tasks did elicit at least some more complex mental state reasoning.

Although the reviewed studies have used different paradigms, ToM-related activity has been reported relatively consistently in several regions, particularly the medial prefrontal cortex (mPFC) and the temporoparietal junction (TPJ). Nevertheless, there is some variation in the patterns of evoked activity and also evidence that there may be at least partially dissociable subsets of regions, for example, for the representation of the mental states of similar and dissimilar others. However there do not appear to be any differences in findings that are consistently related to the paradigm employed.

Interactive Paradigms

Paradigms that directly involve participants are most likely to activate the ToM processes involved in real-life social interactions. Calarge et al. [2003] asked participants in a PET study to invent and say aloud stories describing imaginary encounters with strangers. The authors suggested that this task required participants to place themselves in scenarios requiring mental state attribution. By comparison with a control condition, in which participants read aloud stories requiring no mental state attribution, the ToM task evoked activity in left medial, superior and inferior frontal regions, the anterior, para-, and retrocingulate—extending bilaterally,—the angular gyrus, temporal pole, and the right cerebellum. Although these findings are consistent with other studies, the adequacy of the control condition is questionable. First, although the instructions for both tasks were written, only the control condition required participants to read throughout the task.

[10]Although others have suggested that mirror neurons are involved in the coding of abstract mental states [e.g. Gallese and Goldman, 1998].

Additionally, the control condition was less demanding and engaging than the ToM condition, as participants were not required to "invent" a story. Furthermore, the generation of a coherent narrative would involve executive functioning not required by passive reading during the control task. Indeed, the authors suggested that activation in the angular gyrus and anterior temporal pole may have reflected the language and memory retrieval components of the ToM task.

A number of studies have addressed the issue of participants' "removal" from more traditional ToM scenarios by developing tasks in which participants directly interact with another "person." Interaction with others is the usual situation in which mental states are attributed naturally; successful social interactions rely on these processes occurring rapidly and "on-line." To assess this naturalistic form of ToM reasoning, McCabe et al. [2001] scanned participants whilst they played two-person decision-making games in which they could either cooperate or compete with human or computer opponents. Participants knew whether their opponent was human or computer, and were told that the computer would follow a fixed probabilistic strategy. Furthermore, the computer's "moves" were played immediately to minimize the participants' tendency to attribute mental states to the computer. Participants who cooperated with their opponent demonstrated increased activity in the mPFC when playing against a human compared with the computer, a differentiation that was not seen when participants did not cooperate with their opponent. McCabe et al. suggested that while cooperation required the evaluation of an opponent's mental states, noncooperation may have reflected a tendency to follow a rule-of-thumb strategy for both the computer and human opponents, negating the ToM component of the games.

Gallagher et al. [2002] used a computerized version of the stone-paper-scissors game to investigate on-line ToM reasoning. Participants played against three different "opponents": a "human" competitor, a computer following a simple rule, and a computer making random choices. Because the "opponent" was always a computer, the only difference between the conditions was the intentional stance adopted by the participants, i.e. whether or not opponents were conceived of as being intentional agents in possession of a ToM. Compared with both the rule-solving and random computer conditions, the "person" condition evoked activity in several frontal regions, including the anterior paracingulate. Rilling et al. [2004] also identified activity in the anterior paracingulate in a study in which participants played interactive games requiring the assessment of cooperative intent in ones partner. The study also highlighted the involvement of the TPJ region of the STS.

Explicit Versus Implict Task Instructions

Whether or not participants are explicitly asked to attend to others' mental states might affect which cognitive strategies are used and thus which regions of the brain are activated. It is generally assumed that ToM is an automatic ability and that explicit instructions are not needed to evoke mental state reasoning. Only one behavioral study, however, has specifically tested whether ToM is an automatic process. Apperly et al. [2006] presented participants with videos of a changed-location false belief task, similar in principal to the Sally-Ann false belief task [Baron-Cohen et al., 1985]. Apperly et al. asked participants to monitor either the location of the object (the reality condition), or to keep track of where the woman thought the object was (the belief condition). In both conditions, participants were asked at apparently random time points about either the object's current location or the woman's belief. Given that participants were slower to respond to belief questions whilst monitoring the object's location, when the woman's beliefs were effectively incidental, and as the same effect was not seen for reality questions posed when participants were tracking the woman's beliefs, the authors concluded that mental state reasoning was not automatic.[11]

The results from Apperley's study emphasize the possibility that the precise instructions given to participants in imaging studies might affect the extent to which they engage in mental state reasoning. If mental state reasoning is not automatic, then a task designed to assess ToM that does not explicitly request that participants attend to mental states might not recruit "ToM regions." Only one other study has directly investigated the differences in activation evoked by explicit and implicit instructions, although no reference was made to the automaticity of ToM [Iacoboni et al., 2005]. In their video study, Iacoboni et al. reported that the inferior frontal region implicated in the attribution of intentions was similarly activated when participants were explicitly requested to attend to intentions as when they passively observed the videos. Consequently, the authors argued that the type of mental state reasoning elicited by the task was indeed automatic. Although this conclusion is in contrast to the conclusions drawn by Apperley et al. [2006] on the basis of behavioral data, it is consistent with the theory posed by this review. It should be noted, however, that there are very few imaging studies that have addressed this issue.

[11]An alternative explanation for these findings is that the need to disengage the participants' attention from the focus of interest, either the object's location or the woman's beliefs, would increase the reaction times to questions about the nonattended focus. It seems likely, however that the salience of the object's real location would facilitate the rapid responses to the reality question seen in the belief condition. Furthermore, the salience of the object's real location would increase the time needed for participants to disengage from the object's actual location to respond to the belief question in the reality condition. Although this interpretation assumes that reality is automatically monitored as well as beliefs, we argue that it is not self evident that the findings from this study rule out automatic ToM reasoning in the absence of explicit instructions to attend to mental states.

Within-Subject Comparison of Paradigm Type

Only three studies performed a within-subject comparison of paradigm type [Gallagher et al., 2000; Gobbini et al., 2007; Kobayashi et al., 2007]. Gobbini et al. [2007] compared the patterns of ToM-related activity evoked by two established ToM tasks: the stories used by Gallagher et al. [2000] and the animations of geometric shapes used by Castelli et al. [2000]. As discussed previously, Gobbini et al. replicated the findings of both Gallagher et al. and Castelli et al. They also reported that the ToM conditions in both tasks evoked activity in the anterior paracingulate regions of the mPFC, although there was minimal overlap within this region. Furthermore, they reported differential recruitment of the pSTS and TPJ by the tasks; while activity was seen in the TPJ for the ToM stories, the animations recruited the pSTS. Gobbini et al. suggested that this dissociation resulted from the type of mental state tapped by the two tasks. They suggested that the TPJ was selectively recruited by the stories because they involved more abstract mental states such as false beliefs, whereas the ToM animations could be understood through the perception of motor goals and intentions. Although Gobbini et al. suggested a functional dissociation between the TPJ and pSTS, their argument is that the dissociation is dependent on the type of mental state tapped by the task rather than the task itself. Furthermore, the two paradigms are very different, most notably in terms of language demands, thus confounding the comparison. The possibility that individual mental states might be associated with distinct brain regions is discussed below.

Both Gallagher et al. [2000] and Kobayashi et al. [2007] compared the patterns of activity evoked by ToM stories and cartoons. Although Gallagher et al. did not directly contrast the activity evoked by the two tasks, interaction analyses revealed several regions with increased activity during the cartoon task compared with the story task, including the right middle frontal gyrus, the precuneus and the cerebellum.[12] By contrast, the ToM stories were associated with a significant increase in the extent of activation in the mPFC compared with the cartoons. The authors suggested that the two patterns of activity may have reflected a difference in the level of ToM reasoning elicited by the two tasks, as they were not equated for difficulty. However, Gallagher et al. did not directly contrast the activity evoked by the two tasks. Kobayashi et al. [2007] did perform the direct comparison of second-order false belief stories and cartoons, reporting that ToM cartoons were associated with more activity in medial and dorsolateral regions of the PFC, the right MTG, left lingual gyrus and inferior regions of the right occipital lobe compared with ToM stories. By contrast, the stories evoked more activity in the left amygdala than the cartoons. Unlike Gobbini et al., Kobayashi et al. investigated the same mental state with both paradigms, leading to a more meaningful comparison. Kobayashi et al. did not attempt to interpret the results from the comparison, however, focusing instead on regions of convergent activity to identify modality-independent ToM regions. Furthermore, there were only 16 participants in the study. To be confident in the findings, it would be necessary to replicate them with an increased sample size. Thus, it is not clear in the current literature how paradigmatic variations affect the regions of the brain associated with ToM.

[12]These regions can not be considered as specific to ToM, however, as they also exhibited increased activation for the non-ToM cartoons compared with the jumbled pictures.

Experimental Paradigms: Summary

The findings from the studies described above are summarized in Figure 1; the studies reporting activation in each of the brain regions defined in Table II are grouped according to the paradigm used, e.g. stories, cartoons etc. It is evident that the mPFC/OFC region was most commonly activated by ToM reasoning, regardless of paradigm-type, with only Iacoboni et al. [2005], Mosconi et al. [2005], and Ciaramidaro et al. [2007] failing to observe activation in this region. There is some evidence for partially dissociable subsets of regions differentially supporting ToM reasoning for similar or dissimilar others, vmPFC and dmPFC, respectively [Mitchell et al., 2005b], or the decoupling from reality required for false belief reasoning [Sommer et al., 2007], but there is not a clear distinction between the activations elicited by each paradigm-type.

DO INDIVIDUAL MENTAL STATES RECRUIT DISTINCT BRAIN REGIONS?

The preceding review has largely treated ToM as one domain, and studies investigating the traditional, nonspecific definition of ToM have been considered along with studies of cooperative behavior [e.g. McCabe et al., 2001] and active deception [e.g. Ganis et al., 2003; Kozel et al., 2004]. Most studies have given little consideration to individual mental states, such as thoughts, beliefs and intentions, although Gobbini et al. [2007] partially addressed this issue; thus specific patterns of activity associated with individual mental states may not have been identified. Although relatively few imaging studies have attempted to isolate single mental states, these are reviewed to determine whether individual mental states recruit distinct brain regions.

False Belief

Although false belief paradigms dominate the behavioral assessment of ToM, only 10 imaging studies have investigated the brain activity evoked by this mental state [Gallagher et al., 2000; Gobbini et al., 2007; Grezes et al., 2004b; Kobayashi et al., 2007; Mitchell, 2008; Sabbagh and Taylor, 2000; Saxe and Kanwisher, 2003; Saxe and Powell, 2006; Sommer et al., 2007; Young and Saxe, 2008], three of which

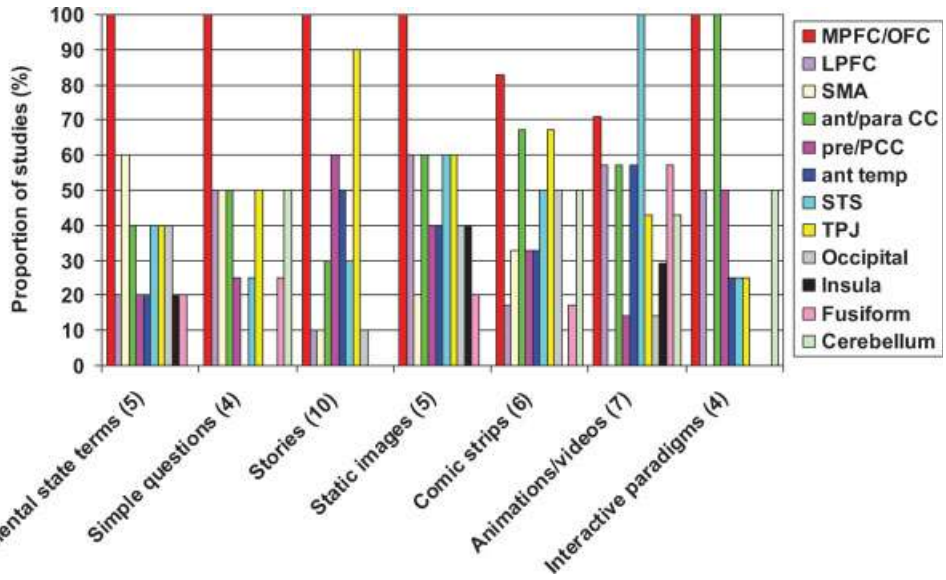

Figure 1.

A comparison of the pattern of brain activity evoked by different ToM tasks. The number of studies implicating each brain region is displayed as a proportion of the number of studies employing each of the following experimental paradigms: mental state terms or word-pairings [Baron-Cohen et al., 1994, 1999; Mason et al., 2004; Mitchell et al., 2005a], simple questions [Ganis et al., 2003; Kozel et al., 2004; Lee et al., 2002; Spence et al., 2001, stories [Fletcher et al., 1995; Gallagher et al., 2000; Gobbini et al., 2007; Kobayashi et al., 2007; Mitchell, 2008; Sabbagh and Taylor, 2000; Saxe and Kanwisher, 2003; Saxe and Powell, 2005; et al., 2001; Young and Saxe, 2008], cartoons and other static images [Gallagher et al., 2000; Goel et al., 1995; Lissek et al., 2008; Mitchell et al., 2005b, 2006], comic strips, including multi-frame cartoons involving a choice phase [Brunet et al., 2000; Ciaramidaro et al., 2007; Kobayashi et al., 2007; Sommer et al., 2007; Vollm et al., 2006; Walter et al., 2004], animations and videos [Castelli et al., 2000; German et al., 2004; Gobbini et al., 2007; Grezes et al., 2004a,b; Iacoboni et al., 2005; Mosconi et al., 2005], and interactive paradigms [Calarge et al., 2003; Gallagher et al., 2002; McCabe et al., 2001; Rilling et al., 2004]. The numbers in parentheses refer to the number of studies in that category.

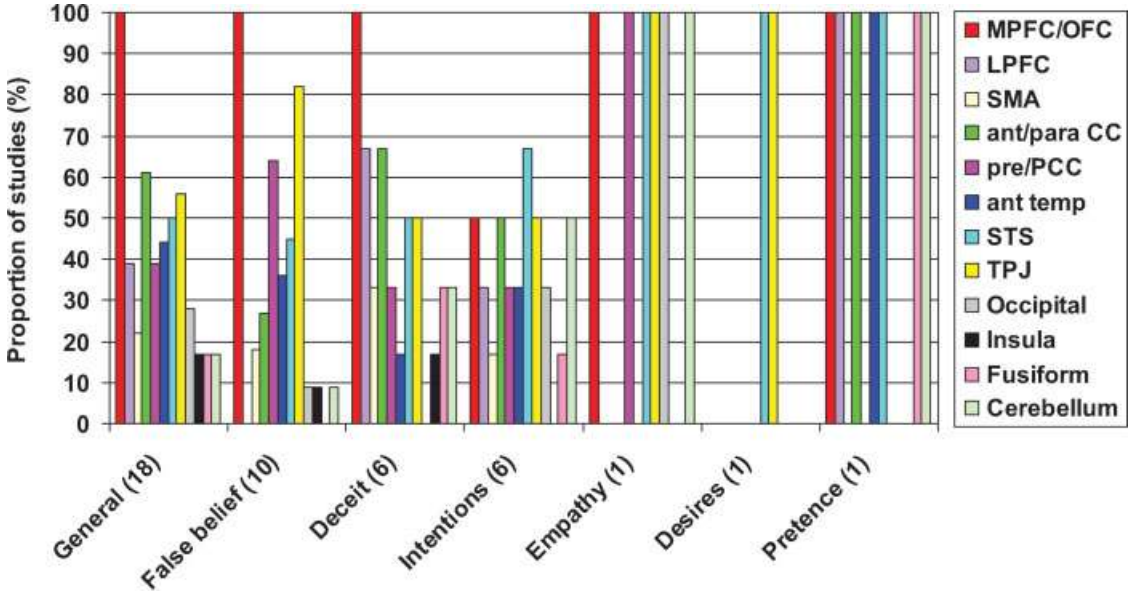

Figure 2.

included the false belief task simply as a ToM "localizer" [Mitchell, 2008; Saxe and Powell, 2006; Young and Saxe, 2008]. These studies all reported activity associated with false belief in the medial prefrontal cortex, although the main frontal focus of ToM-related activity reported by Kobayashi et al. [2007], was slightly more lateral than in the other studies, possibly due to the greater inhibitory control required for attribution of second- rather than first-order false beliefs. Although Sommer et al. [2007] reported activity in the mPFC associated with false belief, they suggested that this activity reflected the cognitive demands inherent in decoupling the representation of the protagonists from reality, rather than the representation of mental states *per se*. Based on the suggestion that the TPJ might be involved in computation of mental states that create perspective differences [Perner et al., 2006], Sommer et al. [2007] suggested that the TPJ, rather than the mPFC, was crucial for false belief reasoning. Interestingly, Kobayashi et al. [2007] also identified the TPJ as one of the two regions associated with false belief, regardless of stimulus modality.

Deception

Several imaging studies have investigated the brain regions activated by deceiving others. Although deception requires the consideration of others' beliefs in the same way that a false belief task might, the act of deceiving another person involves the intentional manipulation of those beliefs [e.g. Ganis et al., 2003]. One paradigm developed to investigate active deception asks participants to answer simple questions. For example, Kozel et al. [2004] asked participants to indicate the correct location of an item in one condition, while in another they were required to indicate the incorrect location. Participants were told that an investigator would be observing their responses to try and detect any deceit. Compared with the truth condition, lying evoked increased activity in regions of the frontal cortex, including the orbitofrontal cortex and frontal gyrus, in the anterior cingulate and in superior temporal and cerebellar regions. The activations in frontal regions were consistent with previous imaging studies in which participants were asked to answer the same questions both truthfully and falsely [Langleben et al., 2002; Spence et al., 2001], although Spence et al. reported a more lateral focus of activation. Furthermore, Spence et al. suggested that the frontal activity reflected executive functions, such as response inhibition rather than the ToM element of deception. Langleben et al. [2002] drew similar conclusions from their study, using a modified version of the Guilty Knowledge Test administered in polygraph interrogation. Consequently, the results from these two studies are not discussed further.

The executive demands inherent in active deception were also emphasized in the studies of Ganis at al. [2003] and Lee et al. [2002]. Although Lee et al. included simple lie and truth conditions, their emphasis was on the regions activated during successful strategies for feigning memory impairments. Bilateral frontal, prefrontal, parietal and temporal activations were reported during the feigned memory task. Although the authors argued that these activations were largely attributable to the executive demands of the task, a nonexecutive role for these regions cannot be ruled out, as the lie vs. truth contrast was not reported. Ganis et al. [2003] investigated activation associated with two different types of lie, arguing that spontaneous lies involve a high degree of executive functioning, because they require working memory, response inhibition, and semantic and episodic retrieval. Memorized lies, however, require only episodic memory retrieval and are consequently less demanding. A general comparison of lies vs. truth revealed activity associated with lying bilaterally in anterior MFC and the fusiform/parahippocampal gyrus, the right precuneus and the left cerebellum. The contrast between the spontaneous and memorized lies, to identify differences attributable to the different executive demands of the conditions, did not reveal activity in the left medial frontal gyrus (MFG), bilateral fusiform/parahippocampal gyrus or the left cerebellum, suggesting that the activity in these areas reflected the mental state processes required for deception.

Lissek et al. [2008] compared the pattern of activity evoked by viewing cartoons of cooperative and/or deceptive interactions between characters. Although both conditions required attribution of mental states to the characters, Lissek et al. hypothesized that the two types of interaction

Figure 2.

A comparison of the brain regions associated with individual mental states. The number of studies implicating each brain region is displayed as a proportion of the number of studies investigating each type of mental state: General [Baron-Cohen et al., 1994, 1999; Calarge et al., 2003; Castelli et al., 2000; Fletcher et al., 1995; Gallagher et al., 2002; Gobbini et al., 2007; Goel et al., 1995; Mason et al., 2004; McCabe et al., 2001; Mitchell et al., 2002, 2005a,b, 2006; Rilling et al., 2004; Saxe and Kanwisher, 2003; Saxe and Powell, 2006; Vogeley et al., 2001], false belief [Gallagher et al., 2000; Gobbini et al., 2007; Grezes et al., 2004b; Kobayashi et al., 2007; Mitchell, 2008; Sabbagh and Taylor, 2000; Saxe and Kanwisher, 2003; Saxe and Powell, 2006; Sommer et al., 2007; Young and Saxe, 2008], deceit [Ganis et al., 2003; Grezes et al., 2004a; Kozel et al., 2004; Lee et al., 2002; Lissek et al., 2008; Spence et al., 2001], intentions [Brunet et al., 2000; Ciaramidaro et al., 2007; Iacoboni et al., 2005; Mosconi et al., 2005; Vollm et al., 2005; Walter et al., 2004], empathy [Vollm et al., 2006], desire [Saxe and Kanwisher, 2003], and pretence [German et al., 2004]. The numbers in parentheses refer to the number of studies in that category.

would be associated with differential activity. The comprehension of cooperation and deception evoked activity in brain regions previously associated with ToM; namely, the TPJ, precuneus, and regions of the posterior cingulate. The perception of deception additionally recruited regions of the PFC, the ACC, and the insula, which the authors attributed to the inherent mismatch between the protagonists' intentions and expectations, and the emotional significance of the deception. Grèzes et al. [2004a] also investigated the brain regions activated during the detection of deceit. Using a video paradigm similar to their false belief study, Grèzes et al. [2004a] observed activity associated with deceit in the anterior temporal lobe, including the amygdala, in addition to medial prefrontal activity previously associated with false belief [Grezes et al., 2004b]. Furthermore the locus of medial prefrontal activation associated with deception was more anterior than the locus reported in the false belief variant of the task. The results from this study, together with the findings from investigations of active deception, largely support the conclusion that although deception and false belief recruit at least partially distinct brain regions, there are some "core" ToM regions, including medial prefrontal regions, the STS and TPJ, that are activated by both mental states.

Intentions and Empathy

As intentions are integral to all purposeful action and guide our behaviors, the recognition and comprehension of others' intentions is essential for normal social understanding. Several studies have investigated the attribution of intentions [Brunet et al., 2000; Ciaramidaro et al., 2007; Iacoboni et al., 2005; Mosconi et al., 2005; Vollm et al., 2006; Walter et al., 2004] and implicated a number of brain regions, including the mPFC, the ACC, and superior temporal regions. Nevertheless, the three studies that have not implicated the mPFC/OFC region in ToM all investigated the attribution of intentions [Ciaramidaro et al., 2007; Iacoboni et al., 2005; Mosconi et al., 2005]. Nevertheless, because the three studies that did implicate the mPFC/OFC region in intentions all employed the same paradigm, it seems probable that paradigmatic differences explain the discrepant findings.

Völlm et al. [2006] included a condition in their comic strip paradigm specifically investigating empathy, the ability to infer and share the emotional states of others. Although empathy specifically refers to emotions, the distinction between emotions and other mental states is somewhat slim. For example, when considering why someone is upset, it is hard to imagine a scenario that does not relate to mental states.[13] The intentions and empathy conditions both evoked activity in regions previously associated with ToM, including medial prefrontal and orbitofrontal regions, the TPJ, and middle and inferior temporal regions [Vollm et al., 2006]. When considered independently, however, empathy was associated with greater activity in medial prefrontal, including the ACC, and amygdalar regions than the attribution of intentions. Conversely, the intentions condition was associated with increased activity more laterally in the frontal cortex and in more superior temporal regions than empathy. These findings are indicative of overlapping yet partially distinct brain regions associated with intentions and empathy.

[13]Indeed, Mitchell et al. [2005a] asked participants to judge the happiness of a face as a measure of ToM.

Summary

The brain regions activated by different mental states are summarized in Figure 2. Clearly individual mental states recruit largely overlapping regions of the brain, although the small number of studies in each group renders any formal comparison impractical. It should also be noted that the findings are confounded by the differences between the paradigms used to investigate each mental state.

VERBAL VERSUS NONVERBAL TASKS

Several authors have suggested that language and verbal interaction may play a crucial role in the development of ToM [e.g. Happe 1995; Marschark 1993; Perner et al., 1994; Peterson and Siegal, 1997, 1998; Yirmiya et al., 1998]. Also syntactical ability is the best predictor of ToM ability in children [Astington and Jenkins, 1999; De Villiers, 1998; Tager-Flusberg and Sullivan, 1994]. Consequently varying linguistic task demands may have contributed to the heterogeneity of the imaging findings.

All three of the studies that employed two different paradigms used one verbal and one nonverbal task [Gallagher et al., 2000; Gobbini et al., 2007; Kobayashi et al., 2007]. Although Gobbini et al. [2007] compared the pattern of activity evoked by ToM stories with 'social' interactions between animated geometric shapes, this was purely qualitative, and the authors did not discuss the possible impact of verbal task demands on the pattern of ToM-related activity. Both Gallagher et al. [2000] and Kobayashi et al., [2007] compared the pattern of activity evoked by ToM stories and cartoons. As previously discussed, Gallagher et al. did not directly contrast the activity evoked by the two tasks, although interaction analyses revealed several regions that were differently activated by the two tasks. Gallagher et al. suggested that different patterns of activity reflected varying levels of task complexity rather than verbal demands. Comparable interaction analyses in the study of Kobayashi et al. revealed increased activity during the ToM story condition compared with the ToM cartoon condition in the left STG and right MTG, but not in the mPFC. The authors suggested that activity in the temporal cortex reflected the increased verbal demands of the story task, based on evidence implicating these regions in language processing.

While the interaction analyses in the studies of Gallagher et al. [2000] and Kobayashi et al. [2007] included effects resulting from both ToM and non-ToM task-components, directly contrasting the two tasks within the ToM condition would yield a less ambiguous view of the affect of verbal factors on the brain regions associated with ToM. Only Kobayashi et al. [2007] performed this contrast. They reported that ToM cartoons evoked increased activity in the right dlPFC, left lingual gyrus and mPFC, the right MTG, and the right inferior occipital gyrus compared with the stories. This finding is in direct contrast to Gallagher et al. [2000], who reported increased medial prefrontal activity during the stories condition. This discrepancy in findings may reflect the differing levels of difficulty of the two tasks; while Gallagher et al. suggested that their story task may have been more demanding than the cartoons, it is possible that the thought bubbles used in Kobayashi et al.'s cartoons may have added complexity and, therefore increased medial prefrontal activity. Furthermore, the activation in the MTG suggests that despite the nonverbal nature of the cartoons, participants may have employed linguistic strategies to complete the task.

Although Gallagher et al. [2000] and Kobayashi et al. [2007] reported some differences in the regions activated by the verbal and nonverbal tasks, both tasks evoked activity in some 'core' ToM regions. Both groups concluded that there were regions of the brain associated with ToM regardless of verbal task demands—the mPFC and the TPJ [Gallagher et al., 2000; Kobayashi et al., 2007]. To investigate this claim, the other reviewed studies were grouped according to their verbal content. Very few of the studies could be considered truly nonverbal, however, with the animations developed by Castelli et al. [2000], Gobbini et al. [2007] and Mosconi et al. [2005] being possible exceptions. Each of these tasks required passive observation only,[14] and the imaging data was acquired during the performance of a completely nonverbal task. In other studies, however, the situation is less clear. For example, in the interactive stone-paper-scissor paradigm [Gallagher et al., 2002], the nonverbal component of the task, deciding which selection to make to beat the opponent, was cued by the visual presentation of "1, 2, 3, GO". Conversely, the task employed by Mitchell et al. [2002, 2006] was essentially verbal, but included pictorial prompts. For the purposes of this comparison, verbal tasks are defined as those in which ToM reasoning was elicited by verbal or numerical stimuli, while tasks incorporating a verbal element only in the cuing or prompt phase have been classed as nonverbal (see Table III). It should be noted, however, that even in apparently entirely nonverbal tasks, the role of linguistic processing can not be ruled out.

Figure 3 illustrates that the mPFC/OFC region was most commonly activated by ToM tasks, regardless of the verbal nature of the paradigm. Nevertheless, this pattern is most obvious in the verbal category, with more than twice as many studies reporting activity in this frontal region as in the next most frequently activated regions, the TPJ and ACC. In the nonverbal category, the second most frequently implicated region was the STS and surrounding cortex. Despite these slight differences in activation, the verbal or nonverbal content of the ToM tasks do not seem to account for the variation in findings between the studies, although it is possible that the slight differences may become more apparent with more studies.

TABLE III. Categorization of verbal and nonverbal studies

Verbal	Nonverbal
Baron-Cohen et al., 1994	Goel et al., 1995
Fletcher et al., 1995	Gallagher et al., 2000
Baron-Cohen et al., 1999	Brunet et al., 2000
Gallagher et al., 2000	Castelli et al., 2000
Sabbagh and Taylor, 2000	Gallagher et al., 2002
McCabe et al., 2001	German et al., 2004
Spence et al., 2001	Grèzes et al., 2004
Vogeley et al., 2001	Grèzes et al., 2004
Lee et al., 2002	Walter et al., 2004
Mitchell et al., 2002	Iacoboni et al., 2005
Calarge et al., 2003	Mitchell et al., 2005b
Ganis et al., 2003	Mosconi et al., 2005
Saxe and Kanwisher, 2003	Völlm et al., 2005
Mason et al., 2004	Ciaramidaro et al., 2007
Rilling et al., 2004	Gobbini et al., 2007
Kozel et al., 2004	Kobayashi et al., 2007
Mitchell et al., 2005a	Sommers et al., 2007
Saxe and Powell, 2005	Lissek et al., 2008
Mitchell et al., 2006	
Gobbini et al., 2007	
Mitchell, 2008	
Kobayashi et al., 2007	
Young and Saxe, 2008	

Verbal paradigms were defined as those that relied heavily on the use of language, such as the tasks involving the comprehension of stories or the recognition of mental state terms. Tasks were not classed as verbal if the only linguistic element of the task was in the prompts used to cue a response.

DISCUSSION

The aim of this review was to determine whether paradigmatic differences, in terms of task-type and mental state(s), could account for the heterogeneous results of imaging studies of ToM, which implicate distinct brain regions in both hemispheres. Neither paradigm type nor the verbal or nonverbal nature of the tasks significantly affected the pattern of ToM-related activity. Nevertheless, there is preliminary evidence that activity in at least partially distinct brain regions may be associated with individual mental states, such as false belief and the detection of deceit [e.g. Grezes et al., 2004a,b]. Also, dissociable patterns of activity have been reported when representing the mental states of similar and dissimilar others [Mitchell

[14]Although in the Castelli task participants were asked to describe what was happening in the animation after each scan.

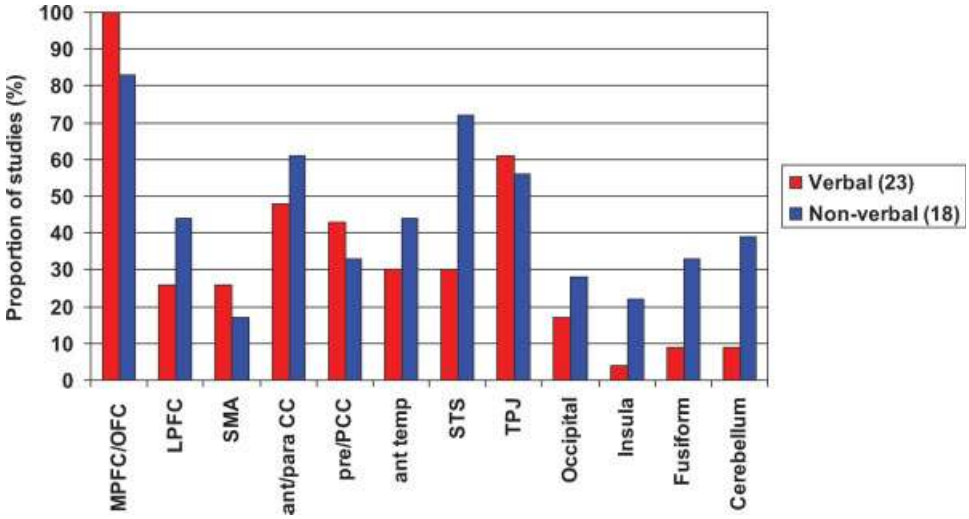

Figure 3.

A comparison of the proportion of verbal and nonverbal tasks that implicate each brain region in ToM. The medial prefrontal and orbitofrontal regions are the most consistently activated region, regardless of whether or not the task is verbal. The results for the verbal and nonverbal tasks are included separately for Gallagher et al. [2000], Gobbini et al. [2007], and Kobayashi et al. [2007]. The numbers in parentheses refer to the number of studies in that category.

et al., 2005b], pointing to another distinction that might be drawn within the broad concept of ToM.

Complex cognitive functions likely involve activity in multiple brain regions rather than being localized to a single "critical" region. Brothers [1990] suggested that three core regions constitute the primate social brain: the OFC, the STS, and the amygdala. In the reviewed studies, the mPFC/OFC region and the STS emerge as "core" regions activated by ToM tasks, but the amygdala appears to be less consistently activated. The TPJ and anterior- and paracingulate cortices also appear to be "core" ToM regions. No single region, however, is recruited in all neuroimaging studies of ToM. Although imaging and lesion studies of ToM both implicate several brain regions, it seems unlikely that each region functions independently. Indeed, inherent in Brother's theory is the idea that a network of interconnected regions underlies social cognition [Brothers, 1990]. Consistent with the idea that ToM is dependant on interaction between brain regions, there are also two reports of ToM deficits following surgical lesions to white matter pathways [Bach et al., 1998; Happe et al., 2001].

The precise nature of brain networks for complex cognition is not well established. For example, the box-and-arrow model of short-term memory proposed by Baddeley and Hitch [1974] did not relate the individual network components to specific regions of the brain. Indeed, current theories about the organization of complex cognition favor distributed functionality and it seems unlikely that a box-and-arrow model could adequately account for functions such as ToM. There is a growing consensus that cognitive functions are dependent on "large-scale cognitive networks that consist of spatially separate computational components, each with its own set of relative specializations that collaborate extensively…" [Just et al., 1999; p 129]. Just et al. [1999] and later Just and Varma [2007] argued that such networks are dynamic, with relative rather than absolute specialization of individual network components. Not only can a brain region perform multiple cognitive functions, but the same cognitive function could be performed by multiple regions, although the precise implementation of that function may vary from region to region. The subtraction techniques typically employed in functional imaging studies assume that activity evoked by different tasks reflects the same cognitive process. If a single region can perform multiple functions, however, this assumption is called into question; does the activity seen in an area during two different tasks really reflect the same cognitive process? The involvement of a region in a cognitive function should only be considered in the context of the overall pattern of brain activity; that is, within the context of a network. Furthermore, sensitive electrophysiological techniques, such as magnetoencephalography, may be used to directly visualize the coherent neuronal activity within and between defined regions, thus helping to clarify whether brain activity really is similar across different tasks.

Just and Varma [2007] also proposed that the function of cortical areas is limited by resource constraints; that is activity is limited by the availability of resources such as the oxygen and glucose necessary for neuronal—and, therefore, computational—function. Furthermore, they proposed that the interactions between regions were similarly constrained, resulting in additional system-wide constraints. Rather than limiting cortical activity, Just and Varma

suggested that these resource constraints shaped the pattern of neural network activity, forcing networks to adapt. For instance, particularly demanding tasks may recruit less specialized regions, while qualitative changes in task demands—for example, if increasing complexity made greater demands on working memory—might also recruit additional, less specialized areas. Such task-dependent topological variation may go some way toward explaining the heterogeneous findings in the neuroimaging literature of ToM. According to these arguments, mPFC may be consistently activated in studies of ToM because the resource constraints limit the activity of more specialized brain regions. Indeed, Saxe and colleagues [Saxe and Powell, 2006; Saxe and Wexler, 2005] suggested that the mPFC is generally involved in social cognition rather than being activated specifically by ToM. Nevertheless, this line of reasoning is speculative and assumes that the critical ToM condition is more demanding than control conditions, which should differ only in terms of social content. The idea that the mPFC may have a more domain-general social role dictated by task complexity or demands could be tested by varying task demands within a ToM condition.

It seems likely that each region in the ToM network is associated with distinct functions that contribute to the accurate comprehension and representation of mental states, although the precise contribution of each region has not yet been established. Alternatively, the specific pattern of interaction between relatively nonspecialized regions—or, as Just et al. speculated, regions with multiple functions—may be critical for ToM reasoning. To fully characterize the ToM network, it will be necessary to identify the roles of each contributing brain region. For example, anterior and posterior regions of the STS have been associated with the perception and representation of biological motion [e.g. Allison et al., 2000; Bonda et al., 1996; Grossman et al., 2000; Pelphrey et al., 2003; Peuskens et al., 2005; Thompson et al., 2005; Vaina et al., 2001] and the role of the STS in ToM may be in utilizing biological [e.g. Blakemore et al., 2003; Schultz et al., 2004][15] or nonbiological [Gobbini et al., 2007] motion cues to understand the protagonists' mental state. Indeed, all the ToM studies that used animation- or video-based paradigms detected activity in the pSTS [Castelli et al., 2000; German et al., 2004; Gobbini et al., 2007; Grezes et al., 2004a,b; Iacoboni et al., 2005; Mosconi et al., 2005]. By contrast, the studies reporting STS activity in the absence of motion, or implied motion, cues reported a more anterior focus of activity [Baron-Cohen et al., 1999; Goel et al., 1995; Lee et al., 2002; Mitchell et al., 2002, 2005b; Rilling et al., 2004; Saxe and Kanwisher, 2003; Saxe and Powell, 2006; Vollm et al., 2006; Young and Saxe, 2008], suggesting possible dissociation of function within the STS. Other studies have implicated the TPJ and additional parietal regions in the attribution of agency to others [e.g. Farrer et al., 2003; Farrer and Frith, 2002], arguably a key process in mental state reasoning [Baron-Cohen, 1995]. Furthermore, the anterior cingulate cortex has been linked to action monitoring and the detection of errors [e.g. Bush et al., 2000], which may be related to the "decoupling" from reality necessary for pretence [Leslie, 1987] and the representation of false beliefs [Sommer et al., 2007].

The brain basis of ToM is likely to be further clarified by functional imaging studies of individuals with ToM deficits. For example, individuals with ASD typically exhibit social impairments in every day life, even though they may perform well on ToM tasks. Nevertheless, the initial imaging findings in ASD are somewhat contradictory. For example, Happé et al. [1996] found that individuals with Asperger Syndrome failed to activate a region of the left medial frontal cortex whilst reading ToM stories, unlike typically developing controls. Baron-Cohen et al. [1999], on the other hand, did report "typical" frontal-temporal activity in individuals with ASD during the Eyes task, but did not observe amygdalar activity. Castelli et al. [2002] reported more widespread functional abnormalities in ASD. Using animations of geometric shapes [Castelli et al., 2000], Castelli et al. [2002] found reduced activity in all regions associated with ToM in typically developing controls, as well as reduced functional connectivity between the extrastriate cortex and the pSTS/TPJ region in individuals with ASD. Castelli et al. [2002] suggested that reduced interaction between higher-level (i.e. ToM) and lower-level (visual) functions may account for the deficit seen in ASD. Schizophrenia is another neuropsychiatric condition associated with impaired ToM [e.g. Sarfati et al., 1997; Sarfati et al., 1999], and aberrant activity during neuroimaging studies of ToM has been reported in mPFC, the ACC, the TPJ, and other temporal regions [Brune et al., 2008; Brunet et al., 2003; Marjoram et al., 2006; Russell et al., 2000]. Taken together, the evidence from neuroimaging studies of individuals with a ToM deficit are consistent with the roles of the brain regions identified as "core" in typically developing individuals.

Although neuroimaging studies can implicate brain regions in particular cognitive functions, they do not reveal which regions are crucial for that function. Lesion studies can identify critical ToM regions, but there are a number of caveats. Firstly, lesions are rarely discreet, and generally affect several, distinct brain regions. Secondly, lesions affecting one brain region are likely to affect the function of connected regions. Finally, studies of acquired brain injury are typically based on small numbers of participants with heterogeneous pathology. Despite these caveats, several studies have reported acquired ToM deficits following damage to the frontal cortex [Rowe et al., 2001; Stone et al., 1998; Stuss et al., 2001], consistent with the evidence from the reviewed neuroimaging studies. Furthermore, two recent lesions studies support fMRI evidence of functional specificity within the prefrontal cortex

[15]Note that these authors did not suggest that this is the only function of the STS.

by demonstrating selective impairments in emotional aspects of ToM, including empathy and reasoning about the feelings of others, following lesions to vmPFC [Shamay-Tsoory and Aharon-Peretz, 2007; Shamay-Tsoory et al., 2003]. Nevertheless, Bird et al. [2004] found no evidence of impaired ToM in a patient with extensive medial prefrontal damage, and it has been suggested that apparent ToM deficits following prefrontal damage may be attributed to more domain-general executive function deficits [Channon and Crawford, 2000]. Additionally lesion studies have implicated other, relatively discrete regions that were also identified in the neuroimaging literature. For example, Apperley et al. [2004] reported an acquired false belief deficit following damage to the TPJ, with apparently preserved executive function. In a later study, however, the same group suggested that the false belief deficit reported in the earlier study was a consequence of a more domain-general executive function impairment [Apperly et al., 2007]. In addition, early amygdalar damage has been associated with impaired ToM, suggesting that this structure may be involved in the development of ToM [Shaw et al., 2004].

Although there is an impressive body of literature relating to brain activity during ToM tasks, there are also some noticeable gaps. It is generally assumed that the mechanisms underlying ToM are the same in all typically developing individuals; indeed, neuroimaging studies depend on this assumption, as the brain activity data from an apparently homogeneous group of individuals are pooled to identify brain regions activated by a task. Nevertheless, everyday experience indicates that there is a range of ToM ability within the typical population: some individuals consistently show insight into mental states and adapt their behavior accordingly, whereas others show less awareness. It seems likely that these individual variations relate at least in part to differences in network function and the strategies employed when considering mental states; these putative differences may underlie some of the heterogeneity in the reviewed studies. Furthermore, several behavioral studies have reported superior ToM performance in females, both in children [e.g. Carlson and Moses, 2001] and adults [Baron-Cohen et al., 1997; Baron-Cohen et al., 2001; Carroll and Chiew, 2006]. Nevertheless, not a single one of the reviewed imaging studies has investigated sex differences in brain activation during ToM

While a difference between the sexes is relatively simple to investigate,—by performing a simple group comparison—it is not immediately apparent how individual differences might be assessed using neuroimaging techniques such as fMRI and PET. Furthermore, although lesion studies provide information at an individual level, to draw any broader conclusions it would be necessary to compare people with identical lesions. Even in instances of surgical lesions, it is highly unlikely that any two lesions are identical, due individual variations in brain structure and constraints on surgical precision. One way to partially resolve the potential problem of individual differences is to also characterize the ToM "network" temporally using magnetoencephalography (MEG). Unlike fMRI, MEG has high temporal resolution, allowing millisecond-by-millisecond visualization of brain activity. MEG, therefore, would be more sensitive to temporal variations that may be associated with individual behavioral differences. Furthermore, it is possible to report meaningful data from individuals using MEG in a way that is not possible with fMRI.

While it is commonly accepted that there is a clear developmental progression in ToM-related skills, neuroimaging studies have usually ignored the ontogeny of ToM. A clear shift during development in the brain regions activated during face processing has been demonstrated using MEG [Kylliainen et al., 2006], but almost nothing is known about developmental change in the brain basis of ToM Only two of the 38 reviewed ToM studies included children [Kobayashi et al., 2007; Mosconi et al., 2005], only one of which compared the brain activity elicited by ToM in children and adults [Kobayashi et al., 2007]. Kobayashi et al. reported that although the ToM tasks elicited activity in several similar regions in the children and adults, there were also differences between the groups, despite comparable behavioral performance. In children ToM elicited more activity in the right vmPFC, the right STG and temporal poles and in the cuneus than in adults. By comparison, adults showed increased ToM-related activity relative to the children only in the left amygdala. These findings raise the possibility that the children used different strategies from the adults and may also suggest an age-related refinement of the regions recruited by mental state reasoning. Nevertheless, the youngest children participating in Kobayashi's study were 8 years of age, whereas the major developmental stages of ToM occur between the ages of 3 and 7 years [e.g. Baron-Cohen et al., 1985; Kobayashi et al., 2007; Wellman et al., 2001; Wimmer, 1983]. Longitudinal studies from ~3 years of age onwards would give a much clearer picture of the development of the neurobiological underpinnings of ToM.

Investigation of developmental changes in the brain basis of ToM may also clarify the contribution of mirror neurons to ToM function. According to the simulation theory, mirror neurons facilitate the simulation of others' mental states using the same neural mechanisms for experiencing those mental states oneself [Gallese and Goldman, 1998]. In particular, mirror neurons have been associated with the comprehension of action goals [e.g. Fogassi et al., 2005; Iacoboni et al., 2005] and imitation [e.g. Iacoboni et al., 1999; Rizzolatti et al., 2001], both of which are thought to be an important stage in the subsequent development of ToM. Consequently, while mirror neurons and simulation may support the development of ToM and the representation of simple mental states, more complex mental state reasoning associated with mature ToM may require additional cognitive processing.

A developmental approach to ToM may also facilitate investigation of the connections between the individual brain regions in the ToM network. Much of this discussion

has focused on network components, with relatively little consideration of their connections. While two case studies have reported an acquired ToM deficit following white matter lesions [Bach et al., 1998; Happe et al., 2001], and fMRI has suggested that functional connectivity is reduced in individuals with a ToM deficit [Castelli et al., 2002], it has only recently been possible to directly investigate the connections between network components. The advent of diffusion tensor imaging (DTI) allows investigation of the structural integrity of brain white matter. Several studies have reported that age correlates positively with fractional anisotropy [Barnea-Goraly et al., 2005; Ben Bashat et al., 2005; Bonekamp et al., 2007; Eluvathingal et al., 2007; Giorgio et al., 2008; Klingberg et al., 1999; Mukherjee et al., 2001; Neil et al., 1998, 2002; Schmithorst et al., 2002; Snook et al., 2005; Suzuki et al., 2003], suggesting that DTI is indeed sensitive to maturation of connective white matter. Once the individual brain regions in the ToM network have been more definitively established, investigating age-related changes in the integrity of the intermediate white matter tracts will be an important next step.

The contribution of network connections to intact ToM reasoning can also be explored by comparing DTI-derived measures of white matter integrity in individuals with ASD with those from typically developing individuals. Such comparisons have revealed both decreased [Alexander et al., 2007; Keller et al., 2007] and increased [Ben Bashat et al., 2007] fractional anisotropy. Furthermore, Barnea-Goraly et al. [2004] reported decreased fractional anisotropy in white matter regions adjacent to cortical regions implicated in social cognition in individuals with ASD. Barnea-Goraly et al. suggested that white matter disruption may contribute to impaired social cognition in ASD. Barnea-Goraly's conclusions could be further explored using a probabilistic tractography algorithm [Behrens et al., 2003] to target specific white matter tracts associated with brain regions implicated in ToM.

It is clear from this discussion that a great deal remains to be done before the brain basis of ToM is more comprehensively established. Although several "core" brain regions have been identified, the variability within the literature cannot be fully accounted for by the paradigmatic variations considered in this review. Clearly the variables we have used to categories neuroimaging studies of ToM are not the only ones that could be applied. Another possible categorization would be according to whether ToM is represented implicitly or explicitly in the task. Although the nature of the instructions given to participants—i.e. whether mental states are referred to explicitly—has been considered, the nature of representation within the task has not. One problem in reviewing the literature is the size of the anatomical regions that we have defined. The mPFC and STS are relatively large regions of the brain, and subdivisions within these regions may be more clearly associated with specific functions. Consequently, studies reporting activity within the same region, e.g. the mPFC, may not be identifying precisely the same area of the brain. Nevertheless, given the heterogeneity within the literature, considering smaller anatomical subdivisions would not currently clarify our understanding of the brain basis of ToM. One solution to this problem would be to look for functional overlap on a voxel-by-voxel basis, although in practise, this is almost impossible due to potential problems with the registration of individuals' brain imaging data, as well as different image acquisition parameters.

In addition to understanding the ontogeny of ToM, as well as individual and sex differences, it is clear that establishing whether reasoning about specific mental states is associated with distinct patterns of brain activity is an important future goal. These different mental states can probably be best investigated by developing a nonverbal paradigm, to avoid linguistic confounds, in which different mental states are investigated without changing any other experimental variables. Finally, investigating the role of functional and structural connections between brain regions associated with ToM will be essential for a complete understanding of the purported network.

REFERENCES

Alexander AL, Lee JE, Lazar M, Boudos R, DuBray MB, Oakes TR, Miller JN, Lu J, Jeong EK, McMahon WM, Bigler ED, Lainhart JE (2007): Diffusion tensor imaging of the corpus callosum in autism. Neuroimage 34:61–73.

Allison T, Puce A, McCarthy G (2000): Social perception from visual cues: Role of the STS region. Trends Cogn Sci 4:267–278.

Apperly IA, Samson D, Chiavarino C, Humphreys GW (2004): Frontal and temporo-parietal lobe contributions to theory of mind: Neuropsychological evidence from a false-belief task with reduced language and executive demands. J Cogn Neurosci 16:1773–1784.

Apperly IA, Riggs KJ, Simpson A, Chiavarino C, Samson D (2006): Is belief reasoning automatic? Psychol Sci 17:841–844.

Apperly IA, Samson D, Chiavarino C, Bickerton WL, Humphreys GW (2007): Testing the domain-specificity of a theory of mind deficit in brain-injured patients: Evidence for consistent performance on non-verbal, "reality-unknown" false belief and false photograph tasks. Cognition 103:300–321.

Astington JW, Jenkins JM (1999): A longitudinal study of the relation between language and theory-of-mind development. Dev Psychol 35:1311–1320.

Bach L, Davis S, Colvin C, Wijerante C, Happe F, Howard R (1998): A neuropsychological investigation of theory of mind in an elderly lady with frontal leucotomy. Cogn Neuropsychiatry 3:139–159.

Baddeley AD, Hitch G (1974):Working memory. In: Bower GH, editor. The Psychology of Learning and Motivation. Advances in Research and Theory. New York: Academic Press. pp 47–89.

Bailey A, Luthert P, Dean A, Harding B, Janota I, Montgomery M, Rutter M, Lantos P (1998): A clinicopathological study of autism. Brain 121(Pt 5):889–905.

Barnea-Goraly N, Kwon H, Menon V, Eliez S, Lotspeich L, Reiss AL (2004): White matter structure in autism: Preliminary evidence from diffusion tensor imaging. Biol Psychiatry 55:323–326.

Barnea-Goraly N, Menon V, Eckert M, Tamm L, Bammer R, Karchemskiy A, Dant CC, Reiss AL (2005): White matter

development during childhood and adolescence: A cross-sectional diffusion tensor imaging study. Cereb Cortex 15:1848–1854.
Baron-Cohen S (1995): Mindblindness: An Essay on Autism and Theory of Mind. Cambridge, Massachusetts; London, England: The MIT Press.
Baron-Cohen S, Leslie AM, Frith U (1985): Does the autistic child have a "theory of mind"? Cognition 21:37–46.
Baron-Cohen S, Leslie AM, Frith U (1986): Mechanical, behavioural and intentional understanding of picture stories in autistic children. Br J Dev Psychol 4:113–125.
Baron-Cohen S, Ring H, Moriarty J, Schmitz B, Costa D, Ell P (1994): Recognition of mental state terms. Clinical findings in children with autism and a functional neuroimaging study of normal adults. Br J Psychiatry 165:640–649.
Baron-Cohen S, Jolliffe T, Mortimore C, Robertson M (1997): Another advanced test of theory of mind: Evidence from very high functioning adults with autism or asperger syndrome. J Child Psychol Psychiatry 38:813–822.
Baron-Cohen S, Ring HA, Wheelwright S, Bullmore ET, Brammer MJ, Simmons A, Williams SC (1999): Social intelligence in the normal and autistic brain: An fMRI study. Eur J Neurosci 11:1891–1898.
Baron-Cohen S, Wheelwright S, Hill J, Raste Y, Plumb I (2001): The "reading the mind in the eyes" Test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism J Child Psychol Psychiatry 42:241–251.
Behrens TE, Johansen-Berg H, Woolrich MW, Smith SM, Wheeler-Kingshott CA, Boulby PA, Barker GJ, Sillery EL, Sheehan K, Ciccarelli O, Thompson AJ, Brady JM, Matthews PM (2003): Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging. Nat Neurosci 6:750–757.
Ben Bashat D, Ben Sira L, Graif M, Pianka P, Hendler T, Cohen Y, Assaf Y (2005): Normal white matter development from infancy to adulthood: Comparing diffusion tensor and high b value diffusion weighted MR images. J Magn Reson Imaging 21:503–511.
Ben Bashat D, Kronfeld-Duenias V, Zachor DA, Ekstein PM, Hendler T, Tarrasch R, Even A, Levy Y, Ben Sira L (2007): Accelerated maturation of white matter in young children with autism: A high b value DWI study. Neuroimage 37:40–47.
Bird CM, Castelli F, Malik O, Frith U, Husain M (2004): The impact of extensive medial frontal lobe damage on 'theory of mind' and cognition. Brain 127(Pt 4):914–928.
Blakemore SJ, Boyer P, Pachot-Clouard M, Meltzoff A, Segebarth C, Decety J (2003): The detection of contingency and animacy from simple animations in the human brain. Cereb Cortex 13:837–844.
Bonda E, Petrides M, Ostry D, Evans A (1996): Specific involvement of human parietal systems and the amygdala in the perception of biological motion. J Neurosci 16:3737–3744.
Bonekamp D, Nagae LM, Degaonkar M, Matson M, Abdalla WM, Barker PB, Mori S, Horska A (2007): Diffusion tensor imaging in children and adolescents: Reproducibility, hemispheric, and age-related differences. Neuroimage 34:733–742.
Brothers L (1990): The social brain: A project for integrating primate behaviour and neurophysiology in a new domain. Concepts Neurosci 1:27–51.
Brune M, Lissek S, Fuchs N, Witthaus H, Peters S, Nicolas V, Juckel G, Tegenthoff M (2008): An fMRI study of theory of mind in schizophrenic patients with "passivity" symptoms. Neuropsychologia 46:1992–2001.
Brunet E, Sarfati Y, Hardy-Bayle MC, Decety J (2000): A PET investigation of the attribution of intentions with a nonverbal task. Neuroimage 11:157–166.
Brunet E, Sarfati Y, Hardy-Bayle MC, Decety J (2003): Abnormalities of brain function during a nonverbal theory of mind task in schizophrenia. Neuropsychologia 41:1574–1582.
Bush G, Luu P, Posner MI (2000): Cognitive and emotional influences in anterior cingulate cortex. Trends Cogn Sci 4:215–222.
Calarge C, Andreasen NC, O'Leary DS (2003): Visualizing how one brain understands another: A PET study of theory of mind. Am J Psychiatry 160:1954–1964.
Carlson SM, Moses LJ (2001): Individual differences in inhibitory control and children's theory of mind. Child Dev 72:1032–1053.
Carroll JM, Chiew KY (2006): Sex and discipline differences in empathising, systemising and autistic symptomatology: Evidence from a student population. J Autism Dev Disord 36:949–957.
Castelli F, Happe F, Frith U, Frith C (2000): Movement and mind: A functional imaging study of perception and interpretation of complex intentional movement patterns. Neuroimage 12:314–325.
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(Pt 8):1839–1849.
Channon S, Crawford S (2000): The effects of anterior lesions on performance on a story comprehension test: Left anterior impairment on a theory of mind-type task. Neuropsychologia 38:1006–1017.
Ciaramidaro A, Adenzato M, Enrici I, Erk S, Pia L, Bara BG, Walter H (2007): The intentional network: How the brain reads varieties of intentions. Neuropsychologia 45:3105–3113.
De Villiers J (1998): On acquiring structural representations for false complements. In: Hollebrandse B, editor. New Perspectives on Language Acquisition. Amherts, MA: GLSA. pp 125–136.
di Pellegrino G, Fadiga L, Fogassi L, Gallese V, Rizzolatti G (1992): Understanding motor events: A neurophysiological study. Exp Brain Res 91:176–180.
Ehrsson HH, Fagergren A, Jonsson T, Westling G, Johansson RS, Forssberg H (2000): Cortical activity in precision–versus power-grip tasks: An fMRI study. J Neurophysiol 83:528–536.
Eluvathingal TJ, Hasan KM, Kramer L, Fletcher JM, Ewing-Cobbs L (2007): Quantitative diffusion tensor tractography of association and projection fibers in normally developing children and adolescents. Cereb Cortex 17:2760–2768.
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.
Farrer C, Franck N, Georgieff N, Frith CD, Decety J, Jeannerod M (2003): Modulating the experience of agency: A positron emission tomography study. Neuroimage 18:324–333.
Fletcher PC, Happe F, Frith U, Baker SC, Dolan RJ, Frackowiak RS, Frith CD (1995): Other minds in the brain: A functional imaging study of "theory of mind" in story comprehension. Cognition 57:109–128.
Fogassi L, Ferrari PF, Gesierich B, Rozzi S, Chersi F, Rizzolatti G (2005): Parietal lobe: From action organization to intention understanding. Science 308:662–667.
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 (3 Pt 1):814–821.

Gallese V, Goldman A (1998): Mirror neurons and the simulation theory of mind-reading. Trends Cogn Sci 2:493–501.

Ganis G, Kosslyn SM, Stose S, Thompson WL, Yurgelun-Todd DA (2003): Neural correlates of different types of deception: An fMRI investigation. Cereb Cortex 13:830–836.

German TP, Niehaus JL, Roarty MP, Giesbrecht B, Miller MB (2004): Neural correlates of detecting pretense: Automatic engagement of the intentional stance under covert conditions. J Cogn Neurosci 16:1805–1817.

Giorgio A, Watkins KE, Douaud G, James AC, James S, De Stefano N, Matthews PM, Smith SM, Johansen-Berg H (2008): Changes in white matter microstructure during adolescence. Neuroimage 39:52–61.

Gobbini MI, Koralek AC, Bryan RE, Montgomery KJ, Haxby JV (2007): Two takes on the social brain: A comparison of theory of mind tasks. J Cogn Neurosci 19:1803–1814.

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

Grezes J, Frith C, Passingham RE (2004a): Brain mechanisms for inferring deceit in the actions of others. J Neurosci 24:5500–5505.

Grezes J, Frith CD, Passingham RE (2004b): Inferring false beliefs from the actions of oneself and others: An fMRI study. Neuroimage 21:744–750.

Grossman E, Donnelly M, Price R, Pickens D, Morgan V, Neighbor G, Blake R (2000): Brain areas involved in perception of biological motion. J Cogn Neurosci 12:711–720.

Happe F, Malhi GS, Checkley S (2001): Acquired mind-blindness following frontal lobe surgery? A single case study of impaired 'theory of mind' in a patient treated with stereotactic anterior capsulotomy. Neuropsychologia 39:83–90.

Happe FG (1994): An advanced test of theory of mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. J Autism Dev Disord 24:129–154.

Happe FG (1995): The role of age and verbal ability in the theory of mind task performance of subjects with autism. Child Dev 66:843–855.

Heider F, Simmel M (1944): An experimental study of apparent behaviour. Am J Psychology 57:243–259.

Iacoboni M, Woods RP, Brass M, Bekkering H, Mazziotta JC, Rizzolatti G (1999): Cortical mechanisms of human imitation. Science 286:2526–2528.

Iacoboni M, Molnar-Szakacs I, Gallese V, Buccino G, Mazziotta JC, Rizzolatti G (2005): rasping the intentions of others with one's own mirror neuron system. PLoS Biol 3:e79.

Just MA, Carpenter PA, Varma S (1999): Computational modeling of high-level cognition and brain function. Hum Brain Mapp 8:128–136.

Just MA, Varma S (2007): The organization of thinking: What functional brain imaging reveals about the neuroarchitecture of complex cognition. Cogn Affect Behav Neurosci 7:153–191.

Keller TA, Kana RK, Just MA (2007): A developmental study of the structural integrity of white matter in autism. Neuroreport 18:23–27.

Kelley WM, Macrae CN, Wyland CL, Caglar S, Inati S, Heatherton TF (2002): Finding the self? An event-related fMRI study. J Cogn Neurosci 14:785–794.

Klingberg T, Vaidya CJ, Gabrieli JD, Moseley ME, Hedehus M (1999): Myelination and organization of the frontal white matter in children: A diffusion tensor MRI study. Neuroreport 10:2817–2821.

Kobayashi C, Glover GH, Temple E (2007): Children's and adults' neural bases of verbal and nonverbal 'theory of mind'. Neuropsychologia 45:1522–1532.

Koski L, Iacoboni M, Dubeau MC, Woods RP, Mazziotta JC (2003): Modulation of cortical activity during different imitative behaviors. J Neurophysiol 89:460–471.

Kozel FA, Revell LJ, Lorberbaum JP, Shastri A, Elhai JD, Horner MD, Smith A, Nahas Z, Bohning DE, George MS (2004): A pilot study of functional magnetic resonance imaging brain correlates of deception in healthy young men. J Neuropsychiatry Clin Neurosci 16:295–305.

Krams M, Rushworth MF, Deiber MP, Frackowiak RS, Passingham RE (1998): The preparation, execution and suppression of copied movements in the human brain. Exp Brain Res 120:386–398.

Kylliainen A, Braeutigam S, Hietanen JK, Swithenby SJ, Bailey AJ (2006): Face and gaze processing in normally developing children: A magnetoencephalographic study. Eur J Neurosci 23:801–810.

Langleben DD, Schroeder L, Maldjian JA, Gur RC, McDonald S, Ragland JD, O'Brien CP, Childress AR (2002): Brain activity during simulated deception: An event-related functional magnetic resonance study. Neuroimage 15:727–732.

Lee TM, Liu HL, Tan LH, Chan CC, Mahankali S, Feng CM, Hou J, Fox PT, Gao JH (2002): Lie detection by functional magnetic resonance imaging. Hum Brain Mapp 15:157–164.

Leslie AM (1987): Pretense and representation: The origins of "theory of mind". Psychol Rev 94:412–426.

Lissek S, Peters S, Fuchs N, Witthaus H, Nicolas V, Tegenthoff M, Juckel G, Brune M (2008): Cooperation and deception recruit different subsets of the theory-of-mind network. PLoS ONE 3:e2023.

Macrae CN, Moran JM, Heatherton TF, Banfield JF, Kelley WM (2004): Medial prefrontal activity predicts memory for self. Cereb Cortex 14:647–654.

Marjoram D, Job DE, Whalley HC, Gountouna VE, McIntosh AM, Simonotto E, Cunningham-Owens D, Johnstone EC, Lawrie S (2006): A visual joke fMRI investigation into theory of mind and enhanced risk of schizophrenia. Neuroimage 31:1850–1858.

Marschark M (1993): Psychological Development if Deaf Children. New York: Oxford University Press.

Mason MF, Banfield JF, Macrae CN (2004): Thinking about actions: The neural substrates of person knowledge. Cereb Cortex 14:209–214.

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 98:11832–11835.

Mitchell JP (2008): Activity in right temporo-parietal junction is not selective for theory-of-mind. Cereb Cortex 18:262–271.

Mitchell JP, Heatherton TF, Macrae CN (2002): Distinct neural systems subserve person and object knowledge. Proc Natl Acad Sci USA 99:15238–15243.

Mitchell JP, Banaji MR, Macrae CN (2005a): General and specific contributions of the medial prefrontal cortex to knowledge about mental states. Neuroimage 28:757–762.

Mitchell JP, Banaji MR, Macrae CN (2005b): The link between social cognition and self-referential thought in the medial prefrontal cortex. J Cogn Neurosci 17:1306–1315.

Mitchell JP, Macrae CN, Banaji MR (2006): Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. Neuron 50:655–663.

Mosconi MW, Mack PB, McCarthy G, Pelphrey KA (2005): Taking an "intentional stance" on eye-gaze shifts: A functional neuroimaging study of social perception in children. Neuroimage 27:247–252.

Mukherjee P, Miller JH, Shimony JS, Conturo TE, Lee BC, Almli CR, McKinstry RC (2001): Normal brain maturation during childhood: Developmental trends characterized with diffusion-tensor MR imaging. Radiology 221:349–358.

Neil J, Miller J, Mukherjee P, Huppi PS (2002): Diffusion tensor imaging of normal and injured developing human brain—a technical review. NMR Biomed 15:543–552.

Neil JJ, Shiran SI, McKinstry RC, Schefft GL, Snyder AZ, Almli CR, Akbudak E, Aronovitz JA, Miller JP, Lee BC, Conturo TE (1998): Normal brain in human newborns: Apparent diffusion coefficient and diffusion anisotropy measured by using diffusion tensor MR imaging. Radiology 209:57–66.

Pelphrey KA, Mitchell TV, McKeown MJ, Goldstein J, Allison T, McCarthy G (2003): Brain activity evoked by the perception of human walking: Controlling for meaningful coherent motion. J Neurosci 23:6819–6825.

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

Perner J, Frith U, Leslie AM, Leekam SR (1989): Exploration of the autistic child's theory of mind: Knowledge, belief, and communication. Child Dev 60:688–700.

Perner J, Ruffman T, Leekham SR (1994): Theory of mind is contagious: You catch it from your sibs. Child Dev 65:1228–1238.

Peterson C, Siegal M (1997): Psychological, biological, and physical thinking in normal, autistic, and deaf children. In: Wellman H, Inagaki K, editors. The Emergence of Core Domains of Thought. San Fransisco: Jossey-Bass. pp 55–70.

Peterson C, Siegal M (1998): Changing focus on the representational mind: Deaf, autistic, and normal children's concepts of false photos, false drawings, and false beliefs. Br J Dev Psychology 16:301–320.

Peuskens H, Vanrie J, Verfaillie K, Orban GA (2005): Specificity of regions processing biological motion. Eur J Neurosci 21:2864–2875.

Premack D, Woodruff G (1978): Chimpanzee problem-solving: A test for comprehension. Science 202:532–535.

Ramnani N, Miall RC (2004): A system in the human brain for predicting the actions of others. Nat Neurosci 7:85–90.

Rilling JK, Sanfey AG, Aronson JA, Nystrom LE, Cohen JD (2004): The neural correlates of theory of mind within interpersonal interactions. Neuroimage 22:1694–1703.

Rizzolatti G, Fadiga L, Gallese V, Fogassi L (1996): Premotor cortex and the recognition of motor actions. Brain Res Cogn Brain Res 3:131–141.

Rizzolatti G, Fogassi L, Gallese V (2001): Neurophysiological mechanisms underlying the understanding and imitation of action. Nat Rev Neurosci 2:661–670.

Rowe AD, Bullock PR, Polkey CE, Morris RG (2001): "Theory of mind" impairments and their relationship to executive functioning following frontal lobe excisions. Brain 124(Pt 3):600–616.

Russell TA, Rubia K, Bullmore ET, Soni W, Suckling J, Brammer MJ, Simmons A, Williams SC, Sharma T (2000): Exploring the social brain in schizophrenia: Left prefrontal underactivation during mental state attribution. Am J Psychiatry 157:2040–2042.

Sabbagh MA, Taylor M (2000): Neural correlates of theory-of-mind reasoning: An event-related potential study. Psychol Sci 11:46–50.

Sarfati Y, Hardy-Bayle MC, Besche C, Widlocher D (1997): Attribution of intentions to others in people with schizophrenia: A non-verbal exploration with comic strips. Schizophr Res 25:199–209.

Sarfati Y, Hardy-Bayle MC, Brunet E, Widlocher D (1999): Investigating theory of mind in schizophrenia: Influence of verbalization in disorganized and non-disorganized patients. Schizophr Res 37:183–190.

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

Saxe R, Powell LJ (2006): It's the thought that counts: Specific brain regions for one component of theory of mind. Psychol Sci 17:692–699.

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

Schmithorst VJ, Wilke M, Dardzinski BJ, Holland SK (2002): Correlation of white matter diffusivity and anisotropy with age during childhood and adolescence: A cross-sectional diffusion-tensor MR imaging study. Radiology 222:212–218.

Schultz J, Imamizu H, Kawato M, Frith CD (2004): Activation of the human superior temporal gyrus during observation of goal attribution by intentional objects. J Cogn Neurosci 16:1695–1705.

Shamay-Tsoory SG, Aharon-Peretz J (2007): Dissociable prefrontal networks for cognitive and affective theory of mind: a lesion study. Neuropsychologia 45:3054–3067.

Shamay-Tsoory SG, Tomer R, Berger BD, Aharon-Peretz J (2003): Characterization of empathy deficits following prefrontal brain damage: The role of the right ventromedial prefrontal cortex. J Cogn Neurosci 15:324–337.

Shamay-Tsoory SG, Lester H, Chisin R, Israel O, Bar-Shalom R, Peretz A, Tomer R, Tsitrinbaum Z, Aharon-Peretz J (2005): The neural correlates of understanding the other's distress: A positron emission tomography investigation of accurate empathy. Neuroimage 27:468–472.

Shaw P, Lawrence EJ, Radbourne C, Bramham J, Polkey CE, David AS (2004): The impact of early and late damage to the human amygdala on 'theory of mind' reasoning. Brain 127(Pt 7):1535–1548.

Snook L, Paulson LA, Roy D, Phillips L, Beaulieu C (2005): Diffusion tensor imaging of neurodevelopment in children and young adults. Neuroimage 26:1164–1173.

Sommer M, Dohnel K, Sodian B, Meinhardt J, Thoermer C, Hajak G (2007): Neural correlates of true and false belief reasoning. Neuroimage 35:1378–1384.

Spence SA, Farrow TF, Herford AE, Wilkinson ID, Zheng Y, Woodruff PW (2001): Behavioural and functional anatomical correlates of deception in humans. Neuroreport 12:2849–2853.

Stone VE, Baron-Cohen S, Knight RT (1998): Frontal lobe contributions to theory of mind. J Cogn Neurosci 10:640–656.

Stuss DT, Gallup GG Jr, Alexander MP (2001): The frontal lobes are necessary for 'theory of mind'. Brain 124(Pt 2):279–286.

Suzuki Y, Matsuzawa H, Kwee IL, Nakada T (2003): Absolute eigenvalue diffusion tensor analysis for human brain maturation. NMR Biomed 16:257–260.

Tager-Flusberg H, Sullivan K (1994): Predicting and explaining behavior: A comparison of autistic, mentally retarded and normal children. J Child Psychol Psychiatry 35:1059–1075.

Thompson JC, Clarke M, Stewart T, Puce A (2005): Configural processing of biological motion in human superior temporal sulcus. J Neurosci 25:9059–9066.

Vaina LM, Solomon J, Chowdhury S, Sinha P, Belliveau JW (2001): Functional neuroanatomy of biological motion perception in humans. Proc Natl Acad Sci USA 98:11656–11661.

Vogeley K, Bussfeld P, Newen A, Herrmann S, Happe F, Falkai P, Maier W, Shah NJ, Fink GR, Zilles K (2001): Mind reading: Neural mechanisms of theory of mind and self-perspective. Neuroimage 14(1 Pt 1):170–181.

Vollm BA, Taylor AN, Richardson P, Corcoran R, Stirling J, McKie S, Deakin JF, Elliott R (2006): Neuronal correlates of theory of mind and empathy: A functional magnetic resonance imaging study in a nonverbal task. Neuroimage 29:90–98.

Walter H, Adenzato M, Ciaramidaro A, Enrici I, Pia L, Bara BG (2004): Understanding intentions in social interaction: the role of the anterior paracingulate cortex. J Cogn Neurosci 16:1854–1863.

Wellman HM, Cross D, Watson J (2001): Meta-analysis of theory-of-mind development: The truth about false belief. Child Dev 72:655–684.

Wicker B, Michel F, Henaff MA, Decety J (1998): Brain regions involved in the perception of gaze: A PET study. Neuroimage 8:221–227.

Williams JH, Whiten A, Suddendorf T, Perrett DI (2001): Imitation, mirror neurons and autism. Neurosci Biobehav Rev 25:287–295.

Wimmer HP (1983): Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. Cognition 13:103–128.

Yirmiya N, Erel O, Shaked M, Solomonica-Levi D (1998): Meta-analyses comparing theory of mind abilities of individuals with autism, individuals with mental retardation, and normally developing individuals. Psychol Bull 124:283–307.

Young L, Saxe R (2008): The neural basis of belief encoding and integration in moral judgment. Neuroimage 40:1912–1920.

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