

Mentalizing about emotion and its relationship to empathy

Christine I. Hooker,¹ Sara C. Verosky,² Laura T. Germine,¹ Robert T. Knight,² and Mark D'Esposito²

¹Department of Psychology, Harvard University, William James Hall, 33 Kirkland St., Cambridge, MA 02138, and ²University of California at Berkeley, Helen Wills Neuroscience Institute, 132 Barker Hall, Berkeley, CA 94720-3190, USA

Mentalizing involves the ability to predict someone else's behavior based on their belief state. More advanced mentalizing skills involve integrating knowledge about beliefs with knowledge about the emotional impact of those beliefs. Recent research indicates that advanced mentalizing skills may be related to the capacity to empathize with others. However, it is not clear what aspect of mentalizing is most related to empathy. In this study, we used a novel, advanced mentalizing task to identify neural mechanisms involved in predicting a future emotional response based on a belief state. Subjects viewed social scenes in which one character had a False Belief and one character had a True Belief. In the primary condition, subjects were asked to predict what emotion the False Belief Character would feel if they had a full understanding about the situation. We found that neural regions related to both mentalizing and emotion were involved when predicting a future emotional response, including the superior temporal sulcus, medial prefrontal cortex, temporal poles, somatosensory related cortices (SRC), inferior frontal gyrus and thalamus. In addition, greater neural activity in primarily emotion-related regions, including right SRC and bilateral thalamus, when predicting emotional response was significantly correlated with more self-reported empathy. The findings suggest that predicting emotional response involves generating and using internal affective representations and that greater use of these affective representations when trying to understand the emotional experience of others is related to more empathy.

Keywords: theory of mind; mentalizing; simulation; mirror neurons; fMRI; emotion recognition; social functioning; empathy

INTRODUCTION

An important aspect of social intelligence is the ability to predict someone else's response and use that prediction to successfully navigate the social exchange (Humphrey, 1976). Because people respond to situations according to their own knowledge and goals, the ability to predict their behavior requires understanding their mental state at the time, that is, their knowledge, emotions, intentions and goals (Frith and Frith, 2006a). The awareness that other people have a mental state, different from our own, which can explain their behavior is referred to as having 'theory of mind' (TOM) or 'mentalizing' skills (Saxe *et al.*, 2004; Frith and Frith, 2006a,b). Particular neurological disorders, such as autism, schizophrenia and frontotemporal dementia, are characterized by deficits in mentalizing skills which lead to poor interpersonal relationships and compromised quality of life (Baron-Cohen, 1995; Brunet *et al.*, 2003; Snowden *et al.*, 2003; Brunet-Gouet and Decety, 2006). Therefore, it is important to identify how different components of mentalizing contribute to the strength of interpersonal relationships and what neural mechanisms facilitate this process. Here, we investigate whether affective mentalizing

(i.e. mentalizing about someone's emotional state) is related to empathy—a process known to facilitate interpersonal relationships.

Mentalizing incorporates inferring mental and emotional state from multiple sources, including non-verbal cues, such as facial expressions and gaze direction, as well as knowledge about the other person's perspective and beliefs (Baron-Cohen, 1995; Frith and Frith, 2005; Frith and Frith, 2006a). The litmus test of rudimentary mentalizing ability is the false belief task which requires predicting the behavior of a character with a false belief. In the classic example, Sally puts a ball in the basket and leaves the room. Ann takes the ball out of the basket and puts it in the box. When Sally comes back into the room, where will she look for the ball? The correct response, that Sally will look in the basket, requires generating a representation of Sally's belief about the situation (Wimmer and Perner, 1983; Wellman and Cross, 2001). More sophisticated mentalizing tasks, such as the faux pas task, require integrating an understanding of False Belief with an understanding of the emotional impact of beliefs. For example, Amy asks Susan what she is wearing to the party, only to find out, subsequently, that Susan does not know about the party and was not invited. Identifying the faux pas requires understanding that someone unintentionally said or did something they should not have and that this behavior has emotional consequences (Stone *et al.*, 1998; Baron-Cohen *et al.*, 1999). Preliminary evidence suggests that higher level mentalizing skills requiring the integration

Received 12 September 2007; Accepted 11 June 2008

Advance Access publication 8 August 2008

The authors would like to thank Jean Decety and Alison Gopnik for helpful discussion. This work was supported by a Postdoctoral Fellowship from the MIND Institute, University of California at Davis and an NIMH Early Career Award K08 MH71746 to CIH.

Correspondence should be addressed to Christine Hooker, PhD, Department of Psychology, Harvard University, William James Hall, 33 Kirkland St., Cambridge MA 02138, USA. E-mail: chooker@wjh.harvard.edu.

of knowledge about beliefs with knowledge about emotions is related to the capacity to empathize with others (Shamay-Tsoory *et al.*, 2003, 2004). However, it is not clear what components of mentalizing are related to empathy.

Although both mentalizing and empathy require an *understanding* of someone else's mental or emotional state, empathy additionally requires *sharing* the emotional experience of the other person (Decety and Jackson, 2004; Singer, 2006). Most research thus far has focused on emotion understanding, particularly the accurate decoding of emotional state based on observable cues (also referred to as *empathic accuracy*). This research reveals a close connection between observing, understanding and sharing emotional response, such that, for example, observing someone else's sadness causes sadness in the observer, and this internal experience (or 'affective representation') of sadness, enables the observer to understand the other person's emotional experience (Levenson *et al.*, 1990; Chartrand and Bargh, 1999). This phenomenon is closely related to perception–action models of empathy (Preston and de Waal, 2002; Meltzoff and Decety, 2003) and suggests that people simulate the experience of others in order to best understand them (Gallese and Goldman, 1998).

However, there are many occasions in which the other person's emotional response is not observable but instead has to be inferred or imagined. These aspects of 'affective mentalizing' (or 'affective TOM') have not been adequately studied. Nonetheless, it is the ability to predict someone else's emotional response which provides the opportunity to use that affective representation to guide behavior in ways that prevent harm or promote well-being in others. For example, imagining how sad a child *would be* if their birthday was forgotten can provide motivation to act in ways that avoid this affective outcome. One possible mechanism, that we investigate in this study, is that the more vividly the future emotional response is imagined, the more that emotional representation can help motivate prosocial behavior which is a primary, adaptive component of empathy (Batson, 1991; de Vignemont and Singer, 2006).

Mentalizing tasks requiring a representation of belief state involve multiple brain regions, including the superior temporal sulcus (STS), temporoparietal junction (TPJ), temporal poles and medial frontal cortex (MFC), including the posterior rostral (prMFC) and the anterior rostral (arMFC) portions (Frith and Frith, 2005, 2006b; Amodio and Frith, 2006). On the other hand, mentalizing tasks focused specifically on identifying emotional state from observable cues involve emotion-related brain regions including the amygdala, anterior insula, thalamus, inferior frontal gyrus (IFG) and somatosensory related cortices (SRC) [including primary somatosensory cortex (SC I) in the postcentral gyrus, secondary somatosensory cortex (SC II) and the supramarginal gyrus] (Adolphs, 2003, 2006).

The SRC and IFG, in particular, have been proposed as regions that facilitate internal modeling or simulation as a

mechanism for empathy. Evidence suggests that the IFG generates an internal motor representation of observed emotions (Carr *et al.*, 2003) and the SRC generates an internal somatic representation of emotional states (Heberlein and Saxe, 2005). Furthermore, lesions in the right SRC produce emotion recognition deficits (Adolphs *et al.*, 2000).

Current study

We designed a study to isolate neural mechanisms involved in predicting a future emotional response of another person and investigate how activity in these neural mechanisms is related to empathy. Our hypothesis is that activity in neural mechanisms supporting the representation of affect, particularly the SRC and IFG, when predicting a new emotional response of someone else will be related to empathy in every day life.

We created an advanced mentalizing task using static, visual social scenes with multiple story characters. Each character's emotional state depends on their belief concerning the social situation: one character has full knowledge, i.e. a True Belief, regarding what is happening in the scene; the other character has partial knowledge or a misperception, i.e. a False Belief. Subjects performed two tasks on each of these two characters: an Emotion Recognition task (i.e. 'What is this person feeling right now?') and an Emotion Inference task: ('What *would* this person feel if he or she had more information?'). The primary condition of interest is the Emotion Inference question regarding the character with a False Belief. In this case, subjects have to imagine and predict a new emotional response which is different than the current observable state of that character. We assessed empathy using the Interpersonal Reactivity Index (IRI) which is a self-report questionnaire considered to be a stable trait measure of empathy (Davis, 1983, 1996). See Figure 1 for an example of the task.

The two tasks are not designed to induce and then measure emotional experience or empathy in the scanner. Instead, the tasks are designed to put demands on a cognitive process—that is, generating a representation of someone else's affective state—and then identify whether greater use of that process (as measured by fMRI activity in regions responsible for affective representations) is related to the amount of empathy that the subject reports experiencing in their daily life.

MATERIALS AND METHODS

Participants

Twenty healthy, English speaking adults (11 females; mean age 21 years; range 19–26 years) volunteered and were paid for their participation. All subjects gave written, informed consent before participation in accordance with the guidelines of the Committee for Protection of Human Subjects at the University of California, Berkeley. Subjects were screened for MR compatibility, neurological and psychiatric illness.

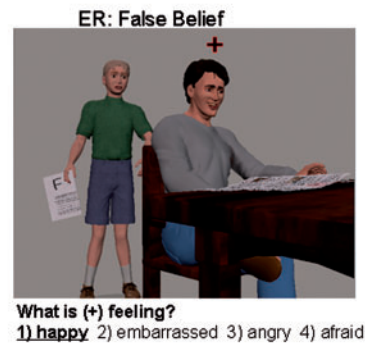
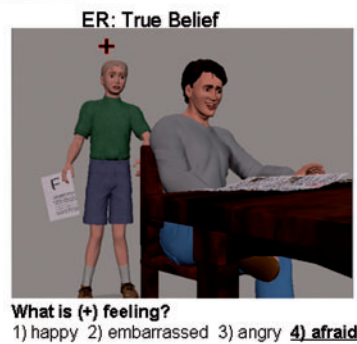
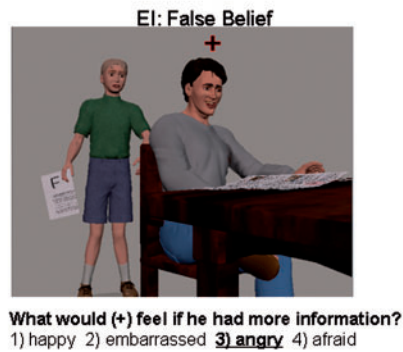
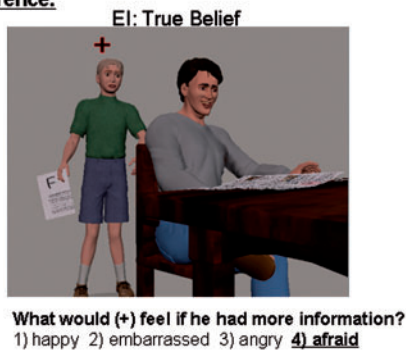
Emotion Recognition:**Emotion Inference:**

Fig. 1 An example of each trial type with the correct response underlined. In the task, the subject is directed to a particular character by the fixation point and is asked an emotion inference or emotion recognition question. In this scene, the son, i.e. the 'True Belief' character, has brought home a failing grade (shown by the 'F' on the paper he is holding), and he appears afraid as he approaches his father. At this point in time, the father, i.e. the 'False Belief' character, has no knowledge of the failing grade, and appears happy.

Task and stimuli

Subjects completed the IRI, which is a 28 item self-report questionnaire assessing empathy and emotional reactivity (Davis, 1996). Subjects read each item and rated on a 5-point scale (0–4) how accurately the statement described them. There are four subscales of the IRI with seven items each. Three of the subscales—empathic concern (EC), perspective taking (PT) and fantasy—are designed to measure different aspects of empathy. The fourth subscale, personal distress, is designed to measure general emotional reactivity (Davis, 1996; Davis *et al.*, 2004).

After filling out the questionnaire, subjects completed a pre-scan comprehension task in which they viewed each of the 40 social scenes and answered factual questions about what was happening in the scene. The comprehension task was designed to familiarize them with the images they would see in the scanner. Subjects examined each scene until they understood the scene and then answered three factual multiple-choice questions, e.g. 'Where is the scene taking place? (i) in a house; (ii) at a carnival; (iii) at an office'.

Subjects were then shown an example trial for each task they would complete in the scanner. For the Emotion Recognition task, subjects were instructed to identify what emotion the character (indicated by a fixation symbol) is feeling at the present time based on available cues. They choose an emotion out of four listed options. In the

Emotion Inference task, subjects are asked what the character (indicated by the fixation symbol) *would* feel if he/she had a full understanding about what is happening in the scene. They are asked these two questions for two characters in each scene in a 2×2 design which yields four trial types: Emotion Recognition: True Belief Character, Emotion Recognition: False Belief Character, Emotion Inference: True Belief Character and Emotion Inference: False Belief Character.

The social scenarios were inspired from mentalizing tasks that currently exist in the literature (Fletcher *et al.*, 1995; Gallagher *et al.*, 2000) but modified in such a way that both emotion judgments could be made of the single visual picture. For example, several scenes involve a character unaware that a positively or negatively valenced event is about to happen (e.g. a terrified mother sees that a car is about to hit her son who is happily riding his bike across the street, or a sad girl is about to open the door to her house and is unaware that a surprise birthday party awaits her). Several scenes involve a misinterpretation (e.g. a wife blames her husband for breaking a vase when the culprit really is a young boy who is hiding from her view). Other scenes involve deception in which one character tricks another (e.g. a man puts an 'I'm stupid' sign on his friend's back but the friend does not see it). In all stimuli, one True Belief Character and one False Belief Character is designated via the

fixation cross. The scenarios include between two and four characters (20% have 2 characters, 65% have 3 characters and 15% have 4 characters).

The four emotion answer choices remain the same for all emotion judgments for a single scene. However, each scene has different emotional facial displays and thus different emotion answer choices. The following eight emotions were represented in the task (in one character or another) and served as answer choices for the stimuli: sad, angry, afraid, happy, amused, annoyed, confused and embarrassed. These emotion choices were based on behavioral pilot testing in which an independent group of subject gave free responses for the Emotion Recognition and Emotion Inference tasks. A positive emotion (happy or amused) was always among the four option choices. Both positive and negative emotions occurred in all trial types, however, specific emotions were not balanced across the different emotion judgment conditions. For our primary condition—the emotion prediction of the False Belief Character—the additional information sometimes causes the False Belief Character to change from a positive to a negative emotion (e.g. the father who is blissfully unaware of his son's failing grade but will be angry when he finds out) and sometimes causes the emotion to change from a negative to positive emotion (e.g. a man believes he is being held-up at gun point, but it is really a friend using the end of a banana to play a trick on him). In total, there are 24 scenes in which the FB character displays a positive emotion (happy, amused) which will change to a negative emotion (sad, angry, afraid, or annoyed) once they have a full understanding. Importantly, because we are contrasting emotion judgments made on the same stimuli within the same task, we control for task demands as well as social and emotional content of the scenes.

Each subject completed four fMRI Emotion Inference/Emotion Recognition scanning runs. The task was designed as a mixed block/event-related design. Each run consisted of 40 trials which were blocked into 20 consecutive trials of each task condition: 20 Emotion Recognition trials and 20 Emotion Inference trials. Within each task block, the specific task trials were presented in a fixed random sequence. Each scene (including the emotion answer choices) was presented for 6 s with a 2, 4 or 6 s jittered inter-trial interval. The sequence of each run was as follows: a rest period (20 s), task instruction indicating Emotion Recognition or Emotion Inference block (2 s), 20 task trials (6 s each with a 2, 4 or 6 s inter-trial interval (ITI)), rest period (20 s), task instruction (2 s), 20 task trials (6 s each with a 2, 4 or 6 s ITI), rest period (20 s).

The presentation sequence of Emotion Recognition and Emotion Inference blocks was counterbalanced across subjects. All pictures were made with the Poser 4 animation program, (Curious Labs, Inc., Scotts Valley, CA, USA). Facial expressions were created by using Facial Action Coding (FACS) algorithms developed for use with the Poser program.

Image acquisition

Images were acquired at 4 Tesla using a Varian INOVA MR scanner (Palo Alto, CA, USA) that was equipped with echo-planar imaging. A standard radiofrequency (RF) head coil was used, and a memory foam pillow restricted head motion. E-Prime software (PST, Pittsburgh, PA, USA) controlled the stimulus display and recorded subject responses. An LCD projector (Epson, Long Beach, CA, USA) projected stimuli onto a backlit projection screen (Stewart, Torrance, CA, USA) within the magnet bore, which the subject viewed via a mirror mounted in the head coil.

Functional images were acquired during four fMRI sessions which began with five dummy scans (with no data acquisition) and four 'blank screen' scans which were subsequently dropped from analysis to insure steady state magnetization for all analyzed data, resulting in 219 whole brain volumes per experimental run, and a total of 876 whole brain volumes for each subject. Images were acquired with parameters used to optimize signal in regions susceptible to drop-out due to magnetic field inhomogeneity. Each volume acquisition included 40, 3.5 mm thick coronal slices with a 0.5 mm inter-slice gap, with a phase encode direction oriented in the superior–inferior direction. A one-shot T2* weighted echo-planar image (EPI) sequence (TR = 2000 ms, TE = 28 ms, FOV = 22.4 cm², matrix size = 64 × 64) was used to acquire blood-oxygenated dependent (BOLD) signal. EPI voxel size at acquisition is 3.5 × 3.5 × 4 mm³. A high-resolution 3D T1-weighted structural scan (MPFLASH sequence) and an in-plane low resolution T2-weighted structural scan (GEMS) were acquired for anatomical localization.

Data processing and analysis

MRI data was processed and analyzed using SPM2 software. Each EPI volume was realigned in space to the first scan, using a six parameter, rigid body, least-squares transformation algorithm. Subjects who showed >3 mm of movement across the session were dropped from analyses. After realignment, we re-sliced the coronal EPI data to the axial plane, and smoothed the data 8 mm (FWHM). We then created and estimated a general linear model (GLM), and created contrast images of the difference between neural activity for each comparison of interest (GLM and data analysis is detailed below). These contrast images were co-registered to the individual subject's co-planar (GEMS) and high resolution (MPFLASH) anatomical images, resliced to 2 × 2 × 2 isotropic voxels and then normalized to the Montreal Neurological Institute (MNI) atlas space.

In the creation of the GLM, the hemodynamic response for each event was modeled from the onset of the trial. We defined each trial type as a covariate of interest: (i) Emotion Inference: False Belief Character; (ii) Emotion Inference: True Belief Character; (iii) Emotion Recognition: False Belief Character; (iv) Emotion Recognition: True Belief Character.

The canonical hemodynamic response function (HRF) was convolved with brain activity at the onset of the trial

type with duration of 4 s. Brain activity was high-pass filtered at 128 s, scaled by the global mean, and corrected for serial autocorrelation. We computed the difference in neural activity between two trial types of interest and then computed whether this difference was significant across subjects by entering the contrast value into a one sample *t*-test. This whole brain random-effects analysis was thresholded at $T(19) = 3.58$, $P < 0.001$ (uncorrected) and a cluster size of 5 voxels for the group activations and $T(12) = 3.93$, $P < 0.001$ with a cluster size of 5 voxels for the correlation analyses. Activity for regions related to *a priori* hypotheses is listed at $P < 0.005$. Cerebellum and brainstem activations are not reported. Only positive activations are shown, unless otherwise specified.

The correlation analysis was performed by entering the contrast value of the comparison of interest and then adding each subject's score on the IRI as a regressor in SPM and identifying regions in a whole brain analysis in which relatively greater activity in the contrast is significantly correlated with higher scores on the IRI. If regions of *a priori* interest showed a significant correlation in the whole brain analysis, contrast values were extracted from the significant cluster to determine whether the data contained statistical outliers (defined as ± 2.5 SD from the mean). Fourteen subjects provided IRI data and were included in the correlation analysis. Correlation coefficients of the IRI and specific regions of interest were determined by conducting bivariate correlation analyses in SPSS.

RESULTS

Behavioral results

Accuracy for the pre-scan comprehension task was 96% (3) with a range of 90–100%, suggesting that subjects understood the social scenes. During the scanning session, subjects performed well on both emotion tasks: Emotion Recognition accuracy = 90% (8) and reaction time (RT) = 3422 ms (416); Emotion Inference accuracy = 77% (9) and RT = 3655 ms (451). A repeated measures 2×2 ANOVA [(Task: EI and ER) \times (Belief: TB and FB)] revealed a main effect of task for accuracy [$F(1,21) = 146$, $P < 0.001$] and RT [$F(1,21) = 26$, $P < 0.001$], such that participants performed more accurately and quicker on the Emotion Recognition task. There was no effect of belief state and no interaction of task by belief state. Mean accuracy and RT are as follows: Emotion Recognition: True Belief: accuracy = 89% (7), RT = 3474 ms (356); Emotion Recognition: False Belief: accuracy = 91% (10), RT = 3381 ms (494); Emotion Inference: True Belief: accuracy = 77% (11), RT = 3686 ms (453); Emotion Inference: False Belief: accuracy = 76% (12), RT = 3650 ms (457).

Imaging results

Planned contrasts were conducted to isolate neural activity when predicting another person's future emotional response. The primary condition is the Emotion Inference with the False Belief Character. The planned contrasts

control for task and emotion related elements of the scene: (i) Emotion Inference: False Belief *vs* Emotion Inference: True Belief controls for the task-related process of *inferring* a future emotional state; (ii) Emotion Inference: False Belief *vs* Emotion Recognition: False Belief controls for directed attention to the False Belief Character, including that character's current emotion; (iii) Emotion Recognition: False Belief *vs* Emotion Recognition: True Belief provides a second control for epiphenomenon related to directed attention to the False Belief Character. Neural regions that are sensitive to current characteristics of the False Belief Character (e.g. conflict detection) should be more active during emotion recognition of this character whereas neural regions that are sensitive to the 'True' emotional content of the scene should be more active for Emotion Recognition of the True Belief Character.

Our prediction is that both mentalizing and emotion-related regions will be active for the emotion prediction, and that greater activity in primarily emotion-related regions will positively correlate with empathy.

Main effect analyses

Emotion Inference: False Belief Character vs Emotion Inference: True Belief Character. In this contrast, Emotion Inference: False Belief $>$ Emotion Inference: True Belief, the task question is the same in both conditions (i.e. "What would this person feel?") and the knowledge state that the subject is considering while answering the question is the same (i.e. if everyone had full knowledge). However, the emotional state will change for the false belief Character whereas it does not for the True Belief Character. As predicted, neural regions involved in both mentalizing and emotion-processing were more active for EI: False Belief than EI: True Belief. Table 1 lists brain regions active in this contrast and Figure 2 displays the data.

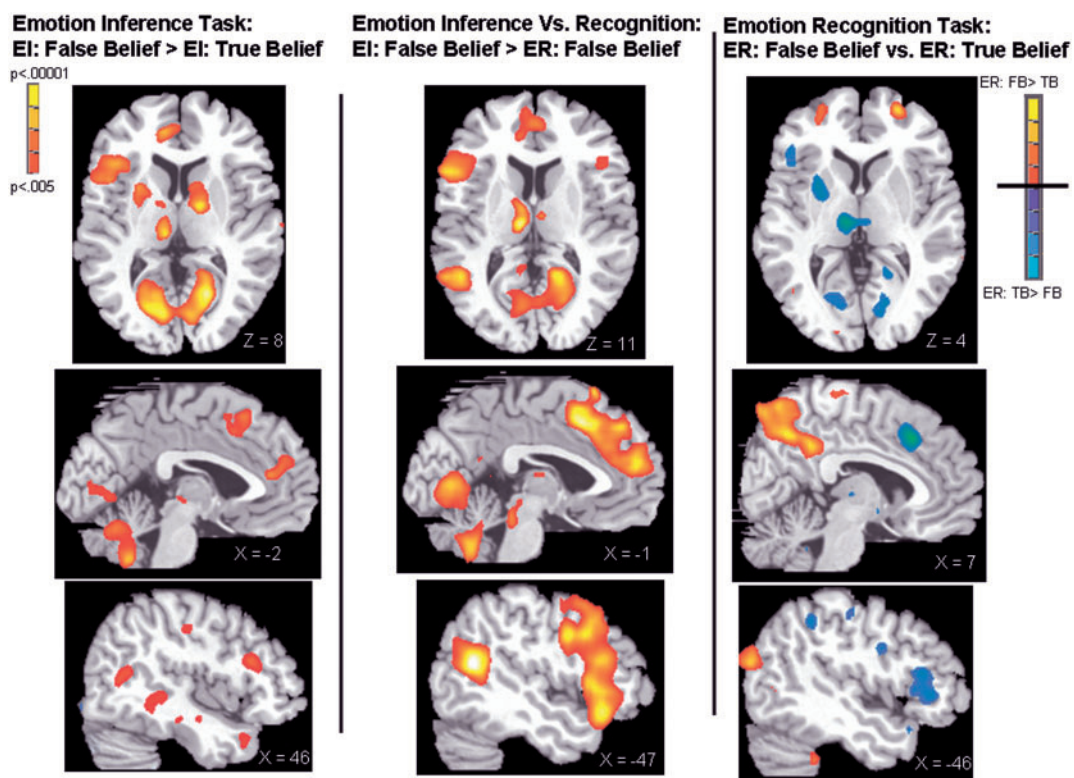
Emotion Inference: False Belief vs Emotion Recognition: False Belief. In the contrast, Emotion Inference: False Belief $>$ Emotion Recognition: False Belief, subjects make a judgment concerning the False Belief Character in both conditions. This controls for activity that might be related to the False Belief Character (e.g. the current emotional expression). There was more activity for EI: FB $>$ ER: FB in mentalizing regions such as the bilateral STS and bilateral MFC with two separate peaks in the arMFC and the prMFC. In addition, there was more activity in emotion-related regions such as the IFG and thalamus. The SRC was not significantly active in this contrast. Table 2 lists brain regions active in this contrast and Figure 2 displays the data.

Interaction of Emotion Inference and Emotion Recognition. We expected a greater increase in SRC and IFG activity when predicting future emotion *vs* the current emotion in the False Belief Character as compared to a relatively smaller increase in activity when predicting future emotion *vs* current emotion in the True Belief Character. This interaction analysis (EI: False Belief–ER: False

Table 1 Neural activity in the Emotion Inference task when subjects imagine what the character with a False Belief would feel if he/she had more knowledge as compared to the same emotion inference question for the True Belief Character

Brain region	Brodmann's area	MNI co-ordinates x, y, z	t -statistic	Volume (voxels)
EI: FB > EI: TB				
L. Inferior frontal gyrus	44, 45	-44, 18, 22	4.5	15
L. Inferior frontal sulcus	46	-38, 32, 22	4.0	194
B. Anterior cingulate cortex (arPFC)	32	-4, 40, 16	4.1	21
B. Supplementary motor area (prPFC)	6	-10, 8, 56	5.8	57
L. Superior temporal sulcus	37	-50, -68, 14	5.2	104
L. Middle temporal gyrus—middle portion	21	-60, -46, -2	5.5	211
L. Inferior temporal gyrus (Temporal pole)	21	-48, 24, -20	3.9	28
L. Lingual—fusiform gyrus	37, 18, 19	-22, -68, -4	8.8	1548
L. Thalamus extending to pallidum		-12, -10, 0	5.3	222
L. Thalamus—pulvinar region		-22, -28, -2	5.2	143
L. Ventral putamen		-20, 6, -10	4.6	11
R. Middle temporal gyrus—middle portion	21	62, -42, -4	4.5	90
R. Thalamus extending to pallidum		14, -2, 10	5.9	321
R. Thalamus—pulvinar region		10, -30, 2	4.2	18
R. Lingual—fusiform gyrus	37, 19	22, -64, 6	7.3	1548
R. Inferior frontal gyrus	44, 45	46, 24, 22	3.5*	
R. Postcentral gyrus (SC I)	3	46, -16, 42	3.0*	

L = Left, R = Right, B = Bilateral.

*Significant at $P < 0.005$.**Fig. 2** Neural activity for each contrast of interest. Data are shown at threshold, $P < 0.005$ to show full extent of activations. The far right panel shows activity for the Emotion Recognition task with significant activity for ER: False Belief > True Belief shown in red colors and ER: True Belief > False Belief shown in blue. All contrasts show IFG, thalamus and striatum activity on the axial slices. Medial prefrontal cortex is shown on the mid-sagittal slice and the somatosensory cortex is shown on the sagittal slices.

Belief) > (EI: True Belief—ER: True Belief) revealed greater activity in bilateral IFG and bilateral thalamus. SC I in the post central gyrus was active at the $P < 0.005$ threshold. Table 3 lists brain regions active in this contrast.

Emotion Recognition: False Belief vs ER: True Belief. There was greater neural activity for the ER: False Belief > ER: True Belief in the right STS/TPJ mentalizing region. The comparison of ER: True Belief vs ER: False

Table 2 Brain regions that were significantly more active for Emotion Inference: False Belief as compared to Emotion Recognition: False Belief

Brain region	Brodmann's area	MNI co-ordinates <i>x, y, z</i>	<i>t</i> -statistic	Volume (voxels)
EI: FB > ER: FB				
L. Precentral gyrus	6	-42, 4, 36	7.7	2861
L. Anterior insula/lateral orbital frontal gyrus	47	-42, 24, -10	6.9	(2861)
L. Inferior frontal gyrus—operculum	44	54, 18, 8	5.6	(2861)
L. Middle frontal gyrus	6	-42, -2, 58	4.9	100
L. Middle frontal gyrus—anterior	9	-18, 50, 38	4.9	38
B. Supplementary motor area/dACC—(prPFC)	8,32, 6	-10, 18, 54	9.7	2321
B. Anterior cingulate cortex (arPFC)	32, 10	2, 56, 18	5.7	(2321)
L. Superior temporal sulcus	21, 39	-54, -58, 16	8.8	1029
L. Middle temporal gyrus—anterior	21	-58, -14, -16	5.2	104
B. Posterior cingulate cortex	23	-12, -54, 30	4.7	84
L. Occipital—lingual gyrus—calcarine	18	-18, -72, -4	7.4	3437
R. Occipital—lingual gyrus—calcarine	17	18, -52, 8	6.5	(3437)
L. Thalamus		-8, -18, 12	7.2	147
L. Ventral pallidum—medial orbital frontal cortex	25	-10, 12, -12	5.5	87
R. Inferior frontal gyrus	44	38, 18, 24	4.0	10
R. Anterior insula	47	32, 22, -6	4.6	59
R. Thalamus		6, -12, 12	4.2	5

L = Left, R = Right, B = Bilateral.

() Parentheses indicates that the volume for that activation is included in the larger cluster listed above.

Table 3 Brain regions that are significantly more active for Emotion Inference: False Belief in the interaction (EI: FB—ER: FB) > (EI: TB—ER: TB)

Brain region	Brodmann's area	MNI co-ordinates <i>x, y, z</i>	<i>t</i> -statistic	Volume (voxels)
(EI: FB—ER: FB) > (EI: TB—ER: TB)				
L. Inferior frontal gyrus—triangular	45	-40, 26, 12	5.0	394
L. Inferior frontal gyrus—operculum	44	-46, 14, 8	4.1	(394)
B. SMA—dorsal anterior cingulate cortex (prPFC)	32	-6, 16, 46	6.6	1118
B. Anterior cingulate cortex (arPFC)	32	-4, 42, 18	5.4	451
L. Middle temporal gyrus—middle portion	37, 21	-52, -46, -8	4.5	113
L. Temporal pole	38	-50, 24, -22	4.2	21
L. Lingual—fusiform gyrus	37, 18, 19	-24, -62, -6	8.8	4043
R. Lingual—fusiform gyrus	37, 19	26, -58, -16	7.2	(4043)
L. Thalamus extending to pallidum		-10, -14, 0	6.8	505
L. Putamen—pallidum		-28, 14, -4	6.5	551
L. Head of the caudate		-12, 2, 14	3.9	12
R. Inferior frontal gyrus	44	42, 26, 18	3.9	19
R. Putamen		14, 14, -2	5.6	299
R. Postcentral gyrus (SC I)	3	48, -16, 48	3.2*	

L = Left, R = Right, B = Bilateral.

() Parentheses indicates that the volume for that activation is included in the larger cluster listed above.

*Significant at $P < 0.005$.

Belief revealed greater activity in emotion-related regions including the left IFG and left thalamus. The post central gyrus was active at $P < 0.005$. Table 4 lists brain regions active in these contrasts and Figure 2 displays the data.

Relationship of neural activity during emotion prediction and self-reported empathy. To identify whether the strength of affective representations during emotion prediction is related to empathy in daily life, we computed the correlation between neural activity during emotion prediction for the False Belief Character (EI: FB) in each main effect contrast with self-reported empathy on the IRI. Our primary hypothesis is that neural activity in the IFG and SRC during emotion prediction (EI: FB > control) will be positively correlated with empathy.

An overall Empathy score was created for each subject by summing the scores from the three empathy-related subscales of the IRI: EC, PT and Fantasy Scale (FS). Correlations between neural activity during specific contrasts and the overall Empathy score were then computed. (Correlations with the IRI subscales showed overlap in the neural regions that correlated with the different empathy subscales, so just the overall Empathy score is reported). The Personal Distress (PD) subscale is a measure of emotionality. High scores on this scale are not associated with prosocial elements of empathy (Davis, 1996; Eisenberg, 2000; Eisenberg *et al.*, 2000). Therefore, this subscale is not included in the overall Empathy score. We report correlations with the PD subscale separately.

Table 4 Neural activity in the Emotion Recognition task

Brain region	Brodmann's area	MNI co-ordinates <i>x, y, z</i>	<i>t</i> -statistic	Volume (voxels)
ER: TB > ER: FB				
L. Inferior frontal gyrus	45	−42, 26, 14	4.0	21
B. Anterior cingulate cortex (arPFC)	32	0, 50, 20	4.0	5
B. Supplementary motor area (prPFC)	32	4, 20, 44	7.6	489
L. Superior frontal gyrus	10	−14, 58, 38	5.1	227
L. Inferior parietal gyrus	40	−52, −40, 48	4.2	19
R. Calcarine	17	20, −52, 6	4.0	10
L. Thalamus		−8, −18, 4	5.9	125
L. Putamen		−28, 6, 4	4.0	53
R. Putamen		14, 12, −4	4.4	20
L. Postcentral gyrus (SC I)	3	−48, −16, 54	3.0*	
ER: FB > ER: TB				
L. Superior frontal gyrus—at frontal pole	11	−26, 52, −4	4.8	121
L. Middle occipital	39, 19	−44, −80, 22	5.0	100
B. Posterior cingulate	23	10, −46, 36	5.7	1661
R. Superior frontal gyrus—superior portion	8	22, 12, 54	6.3	249
R. Superior frontal gyrus—at frontal pole	11	28, 58, 2	5.7	59
R. Superior temporal sulcus—temporoparietal junction	21, 22, 39	50, −54, 22	4.2	378

L = Left, R = Right, B = Bilateral.

*Significant at $P < 0.005$.

There were no outliers in IRI responses. If an *a priori* region of interest was significant in the whole brain analysis, contrast values were extracted from these regions and assessed for outliers.

Neural activity for Emotion Inference: False Belief vs Emotion Inference: True Belief correlated with self-reported empathy. Greater neural activity during EI: False Belief as compared to EI: True Belief in the right SRC and left thalamus was positively correlated with more empathy. The correlation of empathy and right SRC (supramarginal gyrus) = 0.83, $P < 0.01$; empathy and left thalamus = 0.79, $P < 0.01$. Examination of the contrast values revealed one outlier in the right SRC. After this was removed the correlation between empathy and SRC is still significant ($r = 0.75$, $P < 0.01$) (Figure 3). Neural activity in the left IFG was correlated with empathy at a rate just below the cluster threshold ($r = 0.79$, $k = 3$ voxels). The correlation in the right frontal region was centered in the inferior portion of the precentral gyrus (i.e. motor cortex). All correlations are listed in Table 5. These regions did not correlate with PD (Table 6).

Neural activity during Emotion Inference: False Belief vs Emotion Recognition: False Belief correlated with self-reported empathy. Greater neural activity for EI: FB vs ER: FB was positively correlated with empathy in the right supramarginal gyrus, and right thalamus (see Figure 4). After removing one outlier, the correlation of empathy and right supramarginal gyrus was still significant, $r = 0.77$, $P < 0.01$. The correlation between empathy and other regions of interest was as follows: right thalamus, $r = 0.82$, $P < 0.01$; prMPFC, $r = 0.78$, $P < 0.01$. See Table 5 for a full list of brain regions showing a significant correlation. These regions did not correlate with PD.

Neural activity during emotion recognition correlated with self-reported empathy. In the comparison ER: FB > ER: TB, there were no regions that were significantly correlated with empathy. However, greater activity in the left IFG, right somatosensory cortex, right thalamus for ER: True Belief as compared to ER: False Belief was correlated with greater overall Empathy scores. See Table 5 for the full list of regions significantly correlated with overall empathy. We extracted data from the left IFG, right somatosensory cortex, thalamus and amygdala. There were no outliers in any of these regions. The correlation of empathy and the right SC = 0.82, $P < 0.01$, left IFG = 0.85, $P < 0.01$, right thalamus = 0.83, $P < 0.01$ and right amygdala = 0.78, $P < 0.01$ (Figure 5). Brain activity in the thalamus during ER: TB > ER: FB correlated with PD.

DISCUSSION

The aim of this study was to isolate neural regions involved in predicting the emotional response of another person and identifying whether the strength of emotion-related neural activity during this process was related to empathy in daily life. We used a novel, advanced mentalizing task which required predicting emotional response based on belief state. In the main condition, subjects were asked what the character with a false belief would feel if he/she had full knowledge about the situation. In this condition, subjects had to generate the representation of a new emotional state, that is, an emotional state that was different than the emotion displayed on the character's face. In the group analyses, neural regions that are involved in mentalizing, such as the STS/TPJ, MPFC and temporal poles, as well as emotion processing, such as the IFG, SRC and thalamus were more active when predicting a new emotional response for the

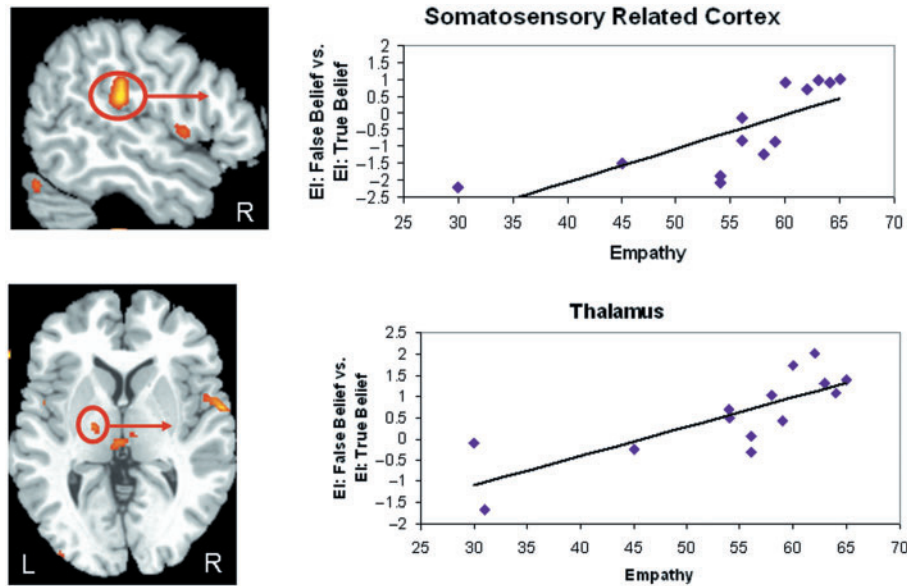


Fig. 3 Greater activity for Emotion Inference: False Belief vs Emotion Inference: True Belief in the right SRC (centered in the supramarginal gyrus) and the left thalamus correlated with empathy.

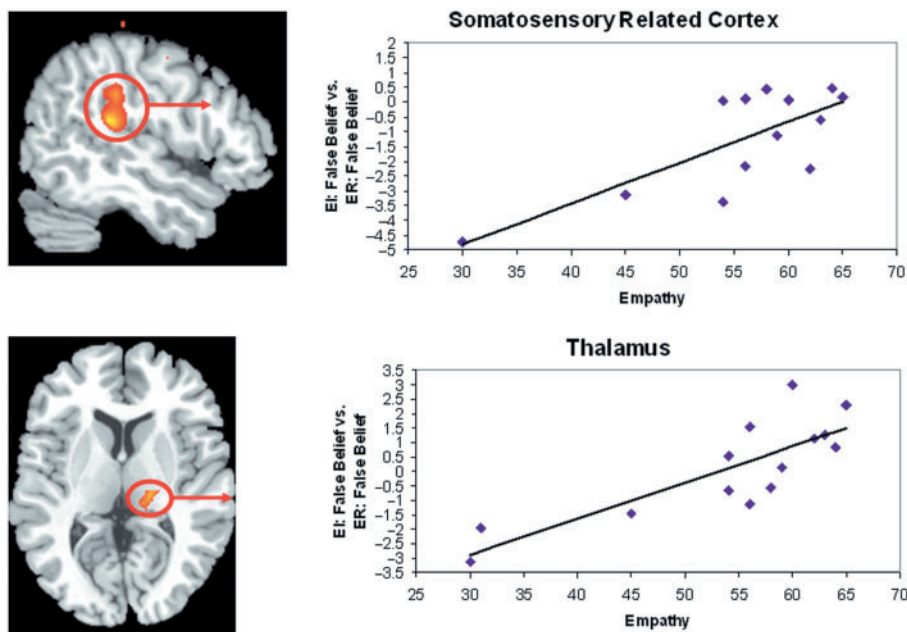


Fig. 4 Greater activity for Emotion Inference: False Belief vs Emotion Recognition: False Belief in right SRC and right thalamus correlated with empathy.

False Belief Character than when recognizing that character's current emotional state or when predicting the emotional response of the True Belief Character whose emotional state was not likely to change. The correlation analysis showed that more activity in primarily emotion-related regions, particularly the right SRC and bilateral thalamus when predicting a new emotional state was related to more self-reported empathy in daily life. In addition, activity in the right SRC, right thalamus and left IFG during emotion recognition of the True Belief Character as compared to the False Belief Character was related to more self-reported empathy.

We interpret our results to suggest that when people predict an emotional response in someone else, they generate an internal affective representation of the predicted emotional response; the stronger the affective representation, the more likely they are to experience empathy in the context of interpersonal relationships. These findings cannot identify the causal relationship between emotion-related activity and empathy. For example, the relationship between neural activity and empathy could indicate that dispositionally empathic people are more practiced at generating emotion representations in the course of their relationships, thus were

Table 5 Brain regions in which enhanced activity for each contrast correlated with self-reported empathy

El:FB > El:TB (Correlated with overall empathy)	BA	MNI Co-ordinates <i>x, y, z</i>	<i>t</i> -Statistic	Volume (voxels)
L. Thalamus		−16, −12, 4	4.3	7
R. Supramarginal gyrus (SRC)	40	56, −26, 22	5.7	88
R. Precentral gyrus (inferior portion)	44	64, 0, 4	5.4	32
B. Posterior cingulate cortex	23	4, −32, 46	4.3	13
B. dACC (prPFC)	24	8, 22, 38	4.6	5
R. Inferior parietal lobe	7	48, −36, 46	4.5	10
R. Thalamus		22, −28, −2	4.4	3ψ
L. Inferior frontal gyrus	44	−46, 4, 4	4.6	3ψ
El:FB vs ER:FB (Correlated with overall empathy)				
B. dACC (prPFC)	24	6, 6, 40	4.4	8
R. Thalamus		20, −22, 6	4.7	43
R. Supramarginal gyrus	40	54, −32, 18	5.5	93
R. Supramarginal gyrus	40	56, −30, 30	4.7	(93)
R. Superior Frontal gyrus	6	26, −2, 72	6.4	99
L. Thalamus		−28, −10, 14	3.3*	
ER:TB > ER:FB (Correlated with overall empathy)				
L. Inferior frontal gyrus	44	−46, 18, 4	5.7	12
L. Fusiform	37	−38, −62, −14	5.1	44
L. Superior parietal lobe	7	−16, −56, 62	4.8	78
R. Postcentral gyrus (SC I)	3	38, −28, 40	5.0	39
R. Supramarginal gyrus—(SRC)	40	48, −24, 22	4.4	11
R. Superior frontal gyrus	6	20, −10, 56	6.2	218
R. Superior parietal lobe	7	52, −36, 60	5.6	14
R. Fusiform	37	32, −58, −20	5.7	70
R. Amygdala	34	28, −8, −14	4.3	10
R. Thalamus		18, −24, 4	4.7	28

L = Left, R = Right, B = Bilateral.

() Parentheses indicates that the volume for that activation is included in the larger cluster listed above.

ψ = below voxel threshold.

*Significant $P < 0.005$.

Table 6 The correlation between neural activity in each main contrast with PD as measured by the IRI

Brain region	Brodmann's area	MNI co-ordinates (<i>x, y, z</i>)	<i>t</i> -statistic	Volume (voxels)
PD: El:FB vs El:TB				
R. Heschl's gyrus/posterior insula	48	54, −18, 12	5.4	38
PD: ER:TB vs ER:FB				
L. Thalamus		−10, −8, 0	4.4	11
L. Calcarine sulcus	19	−8, −74, 14	4.1	5
PD: El:FB vs ER:FB				
R. Superior temporal sulcus	22	42, −50, 8	5.4	17
R. Middle frontal gyrus	9	24, 24, 38	5.4	10

L = Left, R = Right, B = Bilateral.

more likely to use emotion representations in our task. On the other hand, enhanced neural responsivity in emotion processing structures may precipitate and/or increase the likelihood of experiencing empathy in interpersonal

relationships. Though the causal relationship is yet to be determined, the current findings are consistent with the idea that people use emotion-related neural mechanisms to simulate another person's emotional experience in order to best understand their thoughts and feelings (Gallese and Goldman, 1998; Gallese *et al.*, 2004). Furthermore, the findings suggest that more neural activity associated with this internal representation of affect is related to more empathy experienced in daily life.

SRC and thalamus

This simulation model of emotion processing and empathy is best illustrated by data from the right SRC and bilateral thalamus in our task. Prior research has shown that the SRC and the thalamus are related to the perception, experience and recall of emotion (Reiman *et al.*, 1997; Lane *et al.*, 1997b; Damasio *et al.*, 2000; Phan *et al.*, 2002). Across the group of subjects, the SRC was active for predicting the new emotion of the False Belief Character (El:FB > El:TB) as well as for emotion recognition of the True Belief Character ($P < 0.005$, see results and figures). The correlation analysis shows that this neural activity, particularly in the right somatosensory association cortex, was significantly related to individual differences in trait empathy. Specifically, neural activity in this

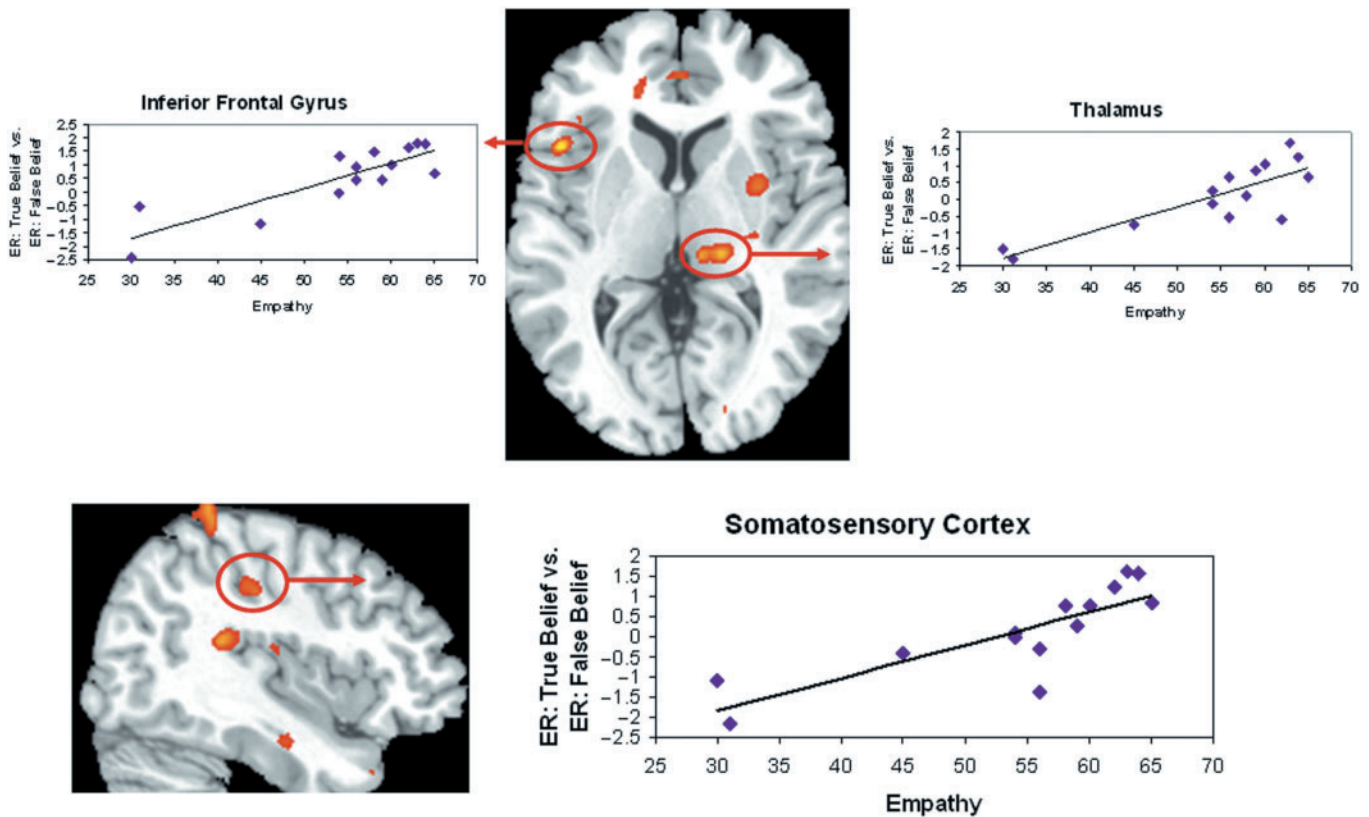


Fig. 5 Greater activity for Emotion Recognition: True Belief vs Emotion Recognition: False Belief in the left IFG, right thalamus and right SRC, with peak in inferior postcentral gyrus, correlated with empathy.

right SRC region during emotion inference of the False Belief Character (EI: FB > EI: TB and ER: FB > ER: FB) as well as during emotion recognition of the True Belief Character (ER: TB > ER: FB) was strongly related to self-reported empathy.

The somatosensory cortices are involved in the perception of touch, temperature and pain as well as the monitoring internal visceral sensations and emotional experiences (Nolte, 1993). It has been suggested that accessing somatic information mediated by the SRC is integral to understanding and empathizing with the experience of others by recreating the observed experience 'as if' it was happening to the self (Damasio *et al.*, 2000). For example, observing painful tactile stimulation applied to the hand of another person activates the hand region of SCI in the observer, and this activity is modulated by the perceived pain intensity for the other person (Avenanti *et al.*, 2005, 2006; Bufalari *et al.*, 2007). The inferior portion of the SRC that is related to empathy in our task has been identified in the literature as integral to emotion processing. Lesions in this region are associated with deficits in facial emotion recognition (Adolphs *et al.*, 2000; Heberlein *et al.*, 2004; Pourtois *et al.*, 2004). Neuroimaging studies show enhanced activity in this region of the SRC when imitating facial emotion (Leslie *et al.*, 2004), identifying emotional state from body motion (Heberlein *et al.*, 2004; Heberlein and Saxe, 2005), and when

thinking about bodily states, such as nausea, dizziness and hunger (Saxe and Powell, 2006). Collectively, these data suggest that subjects in our study may have been accessing a somatic representation of an emotion to make an emotion prediction and that the tendency to do so is related to more empathy experienced in daily life.

Activity in SRC was not related to PD, suggesting that SRC activity when predicting emotion is not related to emotionality, such as panic in stressful situations, but rather is specifically related to the pro-social components of empathy including empathic concern, compassion and perspective-taking (Eisenberg *et al.*, 1995; Eisenberg, 2000).

The thalamus, a region which is broadly involved in emotion processing (Reiman *et al.*, 1997; Phan *et al.*, 2002), was significantly active when predicting a new emotion as compared to every control condition in our task, and enhanced neural activity during emotion prediction in each contrast was significantly correlated with empathy. The thalamus is active during emotion induction studies, which has led to the suggestion that the thalamus is integral to the experience and monitoring of internal feeling states (George *et al.*, 1995; Kimbrell *et al.*, 1999; Damasio *et al.*, 2000). Thalamic activity in our task could reflect accessing a representation of a feeling state in order to predict the emotion for someone else. Neural activity in the thalamus was related to the overall Empathy score, which is a composite of the prosocial

subscales, as well as the PD subscale. This suggests that activity in the thalamus is broadly related to emotionality which may be beneficial for interpersonal relationships but may also be disruptive.

Inferior frontal gyrus

Across the group of subjects, the left IFG was involved in predicting a new emotional state in another person in each contrast (EI: TB > EI: FB and EI: FB > ER: FB) and it was active during emotion recognition of the True Belief Character (ER: TB > ER: FB). Left IFG activity during emotion recognition of the True Belief Character was related to empathy. However, neural activity in the IFG for the other contrasts was not significantly related to empathy. These findings suggest that greater IFG activity in response to current, observable emotional and goal-oriented cues are related to more empathy in daily life. Recent studies are in line with this interpretation. For example, greater IFG activity when listening to another person's actions (i.e. greater mirror-related activity) is related to higher scores on the PT scale of the IRI (Gazzola *et al.*, 2006). Greater activity in the region of the IFG, including the anterior insula, when viewing facial expressions of pleasantness and disgust was related to self-reported empathy on the IRI (Jabbi *et al.*, 2007). Furthermore, psychiatric patients with social functioning deficits, such as those with autism or schizophrenia, have less IFG activity during affective mentalizing tasks (Russell *et al.*, 2000), as well as facial expression imitation (Dapretto *et al.*, 2006), and among autistic patients greater IFG activity during facial expression imitation predicts a higher level of social functioning (Dapretto *et al.*, 2006).

Anterior rostral MFC and posterior rostral MFC

Both the arMFC and the prMFC were involved in predicting emotional response of the False Belief Character as compared to both control conditions (EI: FB > EI: TB and EI: FB > ER: FB) and both MFC regions were more active for emotion recognition of the True Belief Character as compared to the False Belief Character. Furthermore, activity in the prMFC in both emotion prediction contrasts was correlated with empathy.

Because this region is involved in many different cognitive and affective tasks, it is hard to pinpoint the function it is performing when predicting emotional states. Interestingly, although the arMFC is more consistently involved in emotion and mentalizing tasks (Amodio and Frith, 2006), we found that prMFC activity in our task was related to empathy. The prMFC is associated with multiple cognitive tasks, including action monitoring, error monitoring and conflict detection (Amodio and Frith, 2006; van Veen and Carter, 2002, 2006). However, it is also associated with emotion processing, including emotion induction and emotion perception (Phan *et al.*, 2002, 2004). Its function in emotion processing may be to monitor and control emotion-related

arousal (Critchley *et al.*, 2003; Kalisch *et al.*, 2006) or emotional experience (Lane *et al.*, 1997a).

Our results cannot be explained by prMFC activity in response to conflict detection or error monitoring regarding the False Belief Character since we found greater activity in the arMFC and prMFC during emotion recognition of the True Belief as compared to emotion recognition of the False Belief Character. Instead, our results seem more consistent with the idea that during emotion tasks the prMFC is involved in accessing and monitoring internal emotional states, and that this process is related to empathy. For example, the prMFC is active during induced sympathy when hearing about another person's emotional experiences (Decety and Chaminade, 2003; Shamay-Tsoory *et al.*, 2005a). PrMFC activity when observing someone in pain is positively correlated with self-reported empathy on the IRI (Singer *et al.*, 2004). Interestingly, lesion studies have shown that both the prMFC and the arMFC are necessary for mentalizing. For example, it has been shown that prMFC lesion patients have deficits in PT and deception detection (Stuss *et al.*, 2001; Gallagher and Frith, 2003; Stuss and Anderson, 2004; Amodio and Frith, 2006), and arMFC lesion patients have deficits in faux pas detection and these deficits contribute to reduced empathy (Shamay-Tsoory *et al.*, 2003, 2004, 2005b). Our task required perspective taking in order to identify emotion. Future research could try to disentangle emotion from perspective taking to better assess prMFC involvement in emotion processing and empathy.

Simulation vs reasoning based on belief

Interestingly, our results suggest that the right STS/TPJ region does not facilitate mental state understanding through simulation but rather uses a different process. The right STS/TPJ region was particularly responsive to the False Belief Character in all contrasts, including emotion recognition of the False Belief *vs* the True Belief Character, a pattern that was different than the emotion-related network. Furthermore, the right STS/TPJ was not correlated with empathy. This suggests that the right STS/TPJ region—an area more superior and posterior to the STS region associated with biological motion (Allison *et al.*, 2000; Saxe *et al.*, 2004)—may be engaged in reasoning about behavior based on belief state. This interpretation fits with neuroimaging data showing that right STS/TPJ region is more responsive when processing stories or cartoons in which a character with a False Belief is present *vs* absent (Gallagher *et al.*, 2000; Saxe and Kanwisher, 2003), attributing mental states that are incongruent *vs* congruent with the character's background (Saxe and Wexler, 2005), processing facial emotion that is incongruent *vs* congruent with the narrative content of the story (Decety and Chaminade, 2003), and predicting an action motivated by a False Belief *vs* an emotional state (Vollm *et al.*, 2006).

Limitations

There are several limitations to the current study. First, the social scenes contained multiple different emotions and the analyses collapsed across all emotion types. Future research could investigate whether neural activity for a specific emotion (e.g. sad vs happy) is more related to empathy. Second, subjects filled out the empathy questionnaire prior to participating in the fMRI scan. It is not clear whether this influenced neural response to the social scenes. On the other hand, filling out the questionnaire after participating in an emotion judgment experiment may produce biased responding to the questions. It would be helpful to gain a better understanding of how self-report questionnaires may influence task-related neural activity and vice-versa. Finally, the social scenes were created using a graphics program so that specific elements of the scene could be scientifically controlled. However, scenes using real people may produce stronger neural response and thus illustrate more subtle aspects of the relationship between affective mentalizing and empathy.

In summary, the main finding of this study is that activity in emotion-related regions, particularly the right SRC and bilateral thalamus, when predicting a future emotional response of someone else is related to self-reported empathy. This finding suggests that people who generate and use affective information when trying to understand the emotional experience of another person tend to experience more empathy in their interpersonal relationships. This is consistent with the idea that people use simulation to understand others and suggests that using internal affective representations when understanding others is a strategy that may help strengthen interpersonal relationships.

REFERENCES

- Adolphs, R. (2003). Cognitive neuroscience of human social behaviour. *Nature reviews. Neuroscience*, 4, 165–78.
- Adolphs, R. (2006). How do we know the minds of others? Domain-specificity, simulation, and enactive social cognition. *Brain Research*, 1079, 25–35.
- Adolphs, R., Damasio, H., Tranel, D., Cooper, G., Damasio, A.R. (2000). A role for somatosensory cortices in the visual recognition of emotion as revealed by three-dimensional lesion mapping. *The Journal of Neuroscience*, 20, 2683–90.
- Allison, T., Puce, A., McCarthy, G. (2000). Social perception from visual cues: role of the STS region. *Trends in Cognitive Sciences*, 4, 267–78.
- Amodio, D.M., Frith, C.D. (2006). Meeting of minds: the medial frontal cortex and social cognition. *Nature reviews. Neuroscience*, 7, 268–77.
- Avenanti, A., Buetti, D., Galati, G., Aglioti, S.M. (2005). Transcranial magnetic stimulation highlights the sensorimotor side of empathy for pain. *Nature neuroscience*, 8, 955–60.
- Avenanti, A., Paluello, I.M., Bufalari, I., Aglioti, S.M. (2006). Stimulus-driven modulation of motor-evoked potentials during observation of others' pain. *NeuroImage*, 32, 316–24.
- Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. Cambridge, MA: MIT Press.
- Baron-Cohen, S., O'Riordan, M., Stone, V., Jones, R., Plaisted, K. (1999). Recognition of faux pas by normally developing children and children with Asperger syndrome or high-functioning autism. *Journal of Autism and Developmental Disorders*, 29, 407–18.
- Batson, C.D. (1991). *The altruism question: Toward a social-psychological answer*. Hillsdale, N.J.: Erlbaum.
- Brunet-Gouet, E., Decety, J. (2006). Social brain dysfunctions in schizophrenia: a review of neuroimaging studies. *Psychiatry Research*, 148, 75–92.
- Brunet, E., Sarfati, Y., Hardy-Bayle, M.C., Decety, J. (2003). Abnormalities of brain function during a nonverbal theory of mind task in schizophrenia. *Neuropsychologia*, 41, 1574–82.
- Bufalari, I., Aprile, T., Avenanti, A., Di Russo, F., Aglioti, S.M. (2007). Empathy for pain and touch in the human somatosensory cortex. *Cerebral Cortex*, 17, 2553–61.
- Carr, L., Iacoboni, M., Dubeau, M.C., Mazziotta, J.C., Lenzi, G.L. (2003). Neural mechanisms of empathy in humans: a relay from neural systems for imitation to limbic areas. *Proceedings of the National Academy of Sciences of the United States of America*, 100, 5497–502.
- Chartrand, T.L., Bargh, J.A. (1999). The chameleon effect: the perception-behavior link and social interaction. *Journal of Personality Social Psychology*, 76, 893–910.
- Critchley, H.D., Mathias, C.J., Josephs, O., et al. (2003). Human cingulate cortex and autonomic control: converging neuroimaging and clinical evidence. *Brain*, 126, 2139–52.
- Damasio, A.R., Grabowski, T.J., Bechara, A., et al. (2000). Subcortical and cortical brain activity during the feeling of self-generated emotions. *Nature. Neuroscience*, 3, 1049–56.
- Dapretto, M., Davies, M.S., Pfeifer, J.H., et al. (2006). Understanding emotions in others: mirror neuron dysfunction in children with autism spectrum disorders. *Nature. Neuroscience*, 9, 28–30.
- Davis, M.H. (1983). Measuring individual differences in empathy: evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, 44, 113–26.
- Davis, M.H. (1996). *Empathy: a social psychological approach*. Boulder, CO: Westview Press.
- Davis, M.H., Soderlund, T., Cole, J., et al. (2004). Cognitions associated with attempts to empathize: how do we imagine the perspective of another? *Personality and Social Psychology Bulletin*, 30, 1625–35.
- de Vignemont, F., Singer, T. (2006). The empathic brain: how, when and why? *Trends in Cognitive Science*, 10, 435–41.
- Decety, J., Chaminade, T. (2003). Neural correlates of feeling sympathy. *Neuropsychologia*, 41, 127–38.
- Decety, J., Jackson, P.L. (2004). The functional architecture of human empathy. *Behavioral and Cognitive Neuroscience Reviews*, 3, 71–100.
- Eisenberg, N. (2000). Emotion, regulation, and moral development. *Annual Review of Psychology*, 51, 665–97.
- Eisenberg, N., Fabes, R.A., Guthrie, I.K., Reiser, M. (2000). Dispositional emotionality and regulation: their role in predicting quality of social functioning. *Journal of Personality and Social Psychology*, 78, 136–57.
- Eisenberg, N., Fabes, R.A., Murphy, B., Maszk, P., Smith, M., Karbon, M. (1995). The role of emotionality and regulation in children's social functioning: a longitudinal study. *Child Development*, 66, 1360–84.
- Fletcher, P.C., Happe, F., Frith, U., et al. (1995). Other minds in the brain: a functional imaging study of "theory of mind" in story comprehension. *Cognition*, 57, 109–28.
- Frith, C., Frith, U. (2005). Theory of mind. *Current Biology*, 15, R644–6.
- Frith, C.D., Frith, U. (2006a). How we predict what other people are going to do. *Brain Research*, 1079, 36–46.
- Frith, C.D., Frith, U. (2006b). The neural basis of mentalizing. *Neuron*, 50, 531–4.
- Gallagher, H.L., Frith, C.D. (2003). Functional imaging of 'theory of mind'. *Trends in Cognitive Science*, 7, 77–83.
- Gallagher, H.L., Happe, F., Brunswick, N., Fletcher, P.C., Frith, U., Frith, C.D. (2000). Reading the mind in cartoons and stories: an fMRI study of 'theory of mind' in verbal and nonverbal tasks. *Neuropsychologia*, 38, 11–21.
- Gallese, V., Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Science*, 2, 493–501.

- Gallese, V., Keysers, C., Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends in Cognitive Science*, 8, 396–403.
- Gazzola, V., Aziz-Zadeh, L., Keysers, C. (2006). Empathy and the somatotopic auditory mirror system in humans. *Current Biology*, 16, 1824–9.
- George, M.S., Ketter, T.A., Parekh, P.I., Horwitz, B., Herscovitch, P., Post, R.M. (1995). Brain activity during transient sadness and happiness in healthy women. *The American Journal of Psychiatry*, 152, 341–51.
- Heberlein, A.S., Adolphs, R., Tranel, D., Damasio, H. (2004). Cortical regions for judgments of emotions and personality traits from point-light walkers. *Journal of Cognitive Neuroscience*, 16, 1143–58.
- Heberlein, A.S., Saxe, R.R. (2005). Dissociation between emotion and personality judgments: convergent evidence from functional neuroimaging. *NeuroImage*, 28, 770–7.
- Humphrey, N.K. (1976). The social function of intellect. In: Bateson, P.P.G., Hinde, R.A., editors. *Growing points in ethology*. Cambridge, England: Cambridge University Press, pp. 303–17.
- Jabbi, M., Swart, M., Keysers, C. (2007). Empathy for positive and negative emotions in the gustatory cortex. *NeuroImage*, 34, 1744–53.
- Kalisch, R., Wiech, K., Critchley, H.D., Dolan, R.J. (2006). Levels of appraisal: a medial prefrontal role in high-level appraisal of emotional material. *NeuroImage*, 30, 1458–66.
- Kimbrell, T.A., George, M.S., Parekh, P.I., et al. (1999). Regional brain activity during transient self-induced anxiety and anger in healthy adults. *Biological Psychiatry*, 46, 454–65.
- Lane, R.D., Reiman, E.M., Ahern, G.L., Schwartz, G.E., Davidson, R.J. (1997a). Neuroanatomical correlates of happiness, sadness, and disgust. *The American Journal of Psychiatry*, 154, 926–33.
- Lane, R.D., Reiman, E.M., Bradley, M.M., et al. (1997b). Neuroanatomical correlates of pleasant and unpleasant emotion. *Neuropsychologia*, 35, 1437–44.
- Leslie, K.R., Johnson-Frey, S.H., Grafton, S.T. (2004). Functional imaging of face and hand imitation: towards a motor theory of empathy. *NeuroImage*, 21, 601–7.
- Levenson, R.W., Ekman, P., Friesen, W.V. (1990). Voluntary facial action generates emotion-specific autonomic nervous system activity. *Psychophysiology*, 27, 363–84.
- Meltzoff, A.N., Decety, J. (2003). What imitation tells us about social cognition: a rapprochement between developmental psychology and cognitive neuroscience. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences*, 358, 491–500.
- Nolte, J. (1993). *The Human Brain: An introduction to its functional anatomy*, 3rd edn. St. Louis, Missouri: Mosby Year Book, Inc.
- Phan, K.L., Wager, T., Taylor, S.F., Liberzon, I. (2002). Functional neuroanatomy of emotion: a meta-analysis of emotion activation studies in PET and fMRI. *NeuroImage*, 16, 331–48.
- Phan, K.L., Wager, T.D., Taylor, S.F., Liberzon, I. (2004). Functional neuroimaging studies of human emotions. *CNS Spectrums*, 9, 258–66.
- Pourtois, G., Sander, D., Andres, M., et al. (2004). Dissociable roles of the human somatosensory and superior temporal cortices for processing social face signals. *European Journal of Neuroscience*, 20, 3507–15.
- Preston, S.D., de Waal, F.B. (2002). Empathy: its ultimate and proximate bases. *The Behavioral and Brain Science*, 25, 1–20 (Discussion 20–71).
- Reiman, E.M., Lane, R.D., Ahern, G.L., et al. (1997). Neuroanatomical correlates of externally and internally generated human emotion. *The American Journal of Psychiatry*, 154, 918–25.
- Russell, T.A., Rubia, K., Bullmore, E.T., et al. (2000). Exploring the social brain in schizophrenia: left prefrontal underactivation during mental state attribution. *The American Journal of Psychiatry*, 157, 2040–42.
- Saxe, R., Carey, S., Kanwisher, N. (2004). Understanding other minds: linking developmental psychology and functional neuroimaging. *Annual Review of Psychology*, 55, 87–124.
- Saxe, R., Kanwisher, N. (2003). People thinking about thinking people. The role of the temporo-parietal junction in “theory of mind”. *NeuroImage*, 19, 1835–42.
- Saxe, R., Powell, L.J. (2006). It’s the thought that counts: specific brain regions for one component of theory of mind. *Psychological Science*, 17, 692–9.
- Saxe, R., Wexler, A. (2005). Making sense of another mind: the role of the right temporo-parietal junction. *Neuropsychologia*, 43, 1391–9.
- Shamay-Tsoory, S.G., Lester, H., Chisin, R., et al. (2005a). The neural correlates of understanding the other’s distress: a positron emission tomography investigation of accurate empathy. *NeuroImage*, 27, 468–72.
- Shamay-Tsoory, S.G., Tomer, R., Berger, B.D., Aharon-Peretz, J. (2003). Characterization of empathy deficits following prefrontal brain damage: the role of the right ventromedial prefrontal cortex. *Journal of Cognitive Neuroscience*, 15, 324–37.
- Shamay-Tsoory, S.G., Tomer, R., Berger, B.D., Goldsher, D., Aharon-Peretz, J. (2005b). Impaired “affective theory of mind” is associated with right ventromedial prefrontal damage. *The Cognitive and Behavioral Neurology*, 18, 55–67.
- Shamay-Tsoory, S.G., Tomer, R., Goldsher, D., Berger, B.D., Aharon-Peretz, J. (2004). Impairment in cognitive and affective empathy in patients with brain lesions: anatomical and cognitive correlates. *Journal of Clinical and Experimental Neuropsychology*, 26, 1113–27.
- Singer, T. (2006). The neuronal basis and ontogeny of empathy and mind reading: review of literature and implications for future research. *Neuroscience Biobehavioral Reviews*, 30, 855–63.
- Singer, T., Seymour, B., O’Doherty, J., Kaube, H., Dolan, R.J., Frith, C.D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science*, 303, 1157–62.
- Snowden, J.S., Gibbons, Z.C., Blackshaw, A., et al. (2003). Social cognition in frontotemporal dementia and Huntington’s disease. *Neuropsychologia*, 41, 688–701.
- Stone, V.E., Baron-Cohen, S., Knight, R.T. (1998). Frontal lobe contributions to theory of mind. *Journal of Cognitive Neuroscience*, 10, 640–56.
- Stuss, D.T., Anderson, V. (2004). The frontal lobes and theory of mind: developmental concepts from adult focal lesion research. *Brain and Cognition*, 55, 69–83.
- Stuss, D.T., Gallup, G.G.Jr, Alexander, M.P. (2001). The frontal lobes are necessary for ‘theory of mind’. *Brain*, 124, 279–86.
- van Veen, V., Carter, C.S. (2002). The anterior cingulate as a conflict monitor: fMRI and ERP studies. *Physiological Behavior*, 77, 477–82.
- van Veen, V., Carter, C.S. (2006). Error detection, correction, and prevention in the brain: a brief review of data and theories. *Clinical EEG Neuroscience*, 37, 330–5.
- Vollm, B.A., Taylor, A.N., Richardson, P., et al. (2006). Neuronal correlates of theory of mind and empathy: a functional magnetic resonance imaging study in a nonverbal task. *NeuroImage*, 29, 90–8.
- Wellman, H.M., Cross, D. (2001). Theory of mind and conceptual change. *Child Development*, 72, 702–7.
- Wimmer, H., Perner, J. (1983). Beliefs about beliefs: representation and constraining function of wrong beliefs in young children’s understanding of deception. *Cognition*, 13, 103–28.