Social Neuroscience

Matching mind to world and vice versa: Functional dissociations between belief and desire mental state processing

Anna Abraham [a b] , Hannes Rakoczy [c] , Markus Werning [d] , D. Yves von Cramon [a e] & Ricarda
I. Schubotz [e]

[a] Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig

[b] Justus Liebig University, Giessen, Germany

[c] Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

[d] Heinrich-Heine University, Duesseldorf, Germany

[e] Max Planck Institute for Neurological Research, Cologne, Germany

Available online: 10 Aug 2009

PLEASE SCROLL DOWN FOR ARTICLE

Matching mind to world and vice versa: Functional dissociations between belief and desire mental state processing

Anna Abraham

Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, and Justus Liebig University, Giessen, Germany

Hannes Rakoczy

Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

Markus Werning

Heinrich-Heine University, Duesseldorf, Germany

D. Yves von Cramon

Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, and Max Planck Institute for Neurological Research, Cologne, Germany

Ricarda I. Schubotz

Max Planck Institute for Neurological Research, Cologne, Germany

With the aim of understanding how different mental or intentional states are processed in the brain, the present functional magnetic resonance imaging (fMRI) study examined the brain correlates during the ascription of belief intentional states relative to desire intentional states as well as the effect of incongruent relative to congruent intentional states. To this end, sentences containing scenarios were presented to participants and their task was to make judgments concerning the ascription of intentional states based on this information. Belief ascriptions, relative to desire ascriptions, were accompanied by activations in lateral prefrontal structures that include areas known to be involved in relational and conceptual reasoning. Desire ascriptions, in contrast, were accompanied by activations in regions of the medial prefrontal cortex, superior temporal gyri and hippocampal formation, all of which are known for their involvement in self-referential, autobiographical and episodic memory-relevant processes. In addition, the ascription of intentional states that were incongruent with reality (false beliefs and unfulfilled desires) was compared to the ascription of intentional states that were congruent to reality (true belief and fulfilled desires). While no brain region was selectively activated during the processing of unfulfilled desires, the processing of false beliefs was associated with stronger activations in the dorsal medial prefrontal cortex, an area that has been previously linked to the process of decoupling in false belief attribution. These findings provide new insights into more fine-grained aspects of mental state reasoning.

Keywords: Theory of mind; Mentalizing; Functional neuroimaging; Folk psychology; Belief–desire reasoning.

INTRODUCTION

Observing a stranger walk into a bakery can automatically trigger a rapid understanding of the different mental states pertaining to this action, such as the stranger's belief (that she can purchase food in the bakery) and her desire (to satisfy her sweet tooth). Observing the stranger leave the bakery with food would confirm our naive suppositions, whereas seeing her leave the bakery without food but with another person in tow would cause us to revise our original interpretation: Her desire was to meet her friend who she believed would be present in the bakery. This capacity to reason about the unobservable mental states of others and oneself is fundamental to human social behavior and is commonly referred to as theory of mind (Premack & Woodruff, 1978) or mentalizing (Frith & Frith, 2003; Frith, Morton, & Leslie, 1991).

Neuroimaging studies on theory of mind have consistently indicated the involvement of one or more of the following brain regions during mental state reasoning across different experimental paradigms and stimuli (Abraham, Werning, Rakczy, Von Cramon, & Schubotz, 2008; Castelli, Happe, Frith, & Frith, 2000; Fletcher et al., 1995; Gallagher et al., 2000; German, Niehaus, Roarty, Giesbrecht, & Miller, 2004; Perner, Aichhorn, Kronbichler, Staffen, & Ladurner, 2006; Saxe & Powell, 2006; Spiers & Maguire, 2006; Vogeley et al., 2001); the medial prefrontal cortex (mPFC), the temporo-parietal junction (TPJ), the temporal poles, the posterior cingulate and the posterior superior temporal sulcus (Frith & Frith, 2003, 2006; Gallagher & Frith, 2003; Saxe, 2006; Saxe, Carey, & Kanwisher, 2004a). With regard to the functions of the different areas in the network, the temporal poles are held to house social script knowledge (Frith & Frith, 2006; Zahn et al., 2007), whereas the posterior cingulate cortex, an area known for its involvement in episodic memory retrieval and self-referential processing (Vogt & Laureys, 2005; Wagner, Shannon, Kahn, & Buckner, 2005), is possibly involved in accessing such knowledge. The posterior superior temporal sulcus is believed to detect agency, which is plausible given the involvement of this area in spatial perspective taking and responsiveness to biologically relevant movement (Frith & Frith, 2006; Saxe, Xiao, Kovacs, Perrett, & Kanwisher, 2004b). The specific functions of the mPFC and the TPJ are however a matter of debate, with evidence arguing for and against the specific role played by either structure in mental state

attribution (e.g. Abraham et al., 2008; Apperly, Samson, Chiavarino, Bickerton, & Humphreys, 2006b; Ciaramidaro et al., 2007; Decety & Lamm, 2007; Frith & Frith, 2003, 2006; Mitchell, 2008; Perner et al., 2006; Samson, Apperly, Kathirgama-nathan, & Humphreys, 2005; Saxe, 2006; Saxe & Kanwisher, 2003).

Neuroimaging and neuropsychological investigations have been carried out to assess various aspects of mentalizing, including false beliefs vs. false photographs (Saxe & Kanwisher, 2003), thoughts vs. appearances and bodily sensations (Saxe & Powell, 2006), cognitive vs. affective aspects of false belief reasoning and mental inferences (Shamay-Tsoory & Aharon-Peretz, 2007), mental state attribution vs. action descriptions (Castelli et al., 2000), animacy vs. inanimacy judgments (Wheatley, Milleville, & Martin, 2007), mental state appraisals vs. physical appearance appraisals (Mitchell, Banaji & Macrae, 2005), and intentional relational reasoning vs. non-intentional relational reasoning (Abraham et al., 2008). Within the domain of intentional mental states, most studies have mainly investigated reasoning about beliefs, and false beliefs in particular (e.g., Saxe & Kanwisher, 2003), whereas other intentional states are only rarely the focus of study (Saxe & Wexler, 2005).[1]

The present work, therefore, focuses on the neural basis of the ascription of different types of intentional states.[2] Our folk psychology distinguishes between

---

[1] The only mention of an experimental contrast between beliefs and desires within the same paradigm was in a review article by Saxe et al. (2004a, p. 109), where it was reported that the TPJ was significantly more responsive to beliefs than to desires in one previous study (Saxe & Kanwisher, 2003). However, it was also stated within the same context that the stimuli used in this study and others (that investigated the attribution of desires, goals and intentions) were suboptimal because "none of the stimuli in these studies were designed to exclude belief attribution." Saxe (2006) has argued for the role played by the TPJ in underlying the representational aspects of mental states such as beliefs, and has also provided support for a special role played by the right TPJ in desire processing (Saxe & Wexler, 2005).

[2] We introduce the term "intentional states" here to avoid misunderstandings that could stem from the more generalized usage of the term "mental states" in social neuroscience, which is used to refer a wide variety of phenomena under the general umbrella of mentalizing or mental state reasoning. We employ the term "intentional states" to refer to specific mental states such as beliefs and desires that are held, within the purview of philosophy of mind, to characteristically exhibit "intentionality" or the capacity of the mind to be about or to represent things, properties and states of affairs (Dennett & Haugeland, 1987; Perry, 1994).

two broad classes of intentional mental states. First, there are cognitive attitudes, that is, attitudes about what *is* the case. These attitudes have a mind-to-world direction of fit in that they aim toward truth. Beliefs are the paradigm case in this respect (Anscombe, 1957; Searle, 1983). Second, there are conative or "pro" attitudes (Davidson, 1963). These attitudes have a world-to-mind direction of fit in that they aim at changing the world according to their content. Desires are the paradigm case here. While beliefs and desires can have the same content (for example, the belief that it rains and the desire that it rains), they differ fundamentally in their direction of fit and the kind of attitude taken toward the content (taking for true in the case of beliefs vs. valuing in the case of desires). Together, cognitive and conative attitudes constitute reasons for acting and are jointly referred to in rational action explanation, such as in "He bought the stock because he *thought* that would bring profit, and he *wanted* to get rich." Folk psychology is accordingly often called "belief–desire" psychology.

Developmentally, children seem to acquire an understanding of conative attitudes before they understand analogous cognitive ones (e.g., Rakcozy, Warneken, & Tomasello, 2007; Wellman & Bartsch, 1988). So, for instance, two-year-olds understand that different people can want different things but not that they think different things (Repacholi & Gopnik, 1997), and they understand the desire-dependent emotions of happiness/frustration, but do not understand the corresponding belief-dependent emotion of surprise (Hadwin & Perner, 1991).

Against the background of this developmental asymmetry, the central question from the point of view of cognitive neuroscience is whether there are corresponding functional differences in the neural substrates when making belief vs. desire ascriptions. The aim of the present study, therefore, was to systematically investigate the neural bases of understanding cognitive (belief) vs. conative (desire) attitudes. Furthermore, following up on recent work on the ascription of true vs. false beliefs (Sommer et al., 2007), the role of match or mismatch between the content of the propositional attitude and states of affairs in the world for both kinds of intentional states were investigated: reasoning about true beliefs and fulfilled desires (content matches the world/congruent) was compared to reasoning about false beliefs and unfulfilled desires (content does not match the world/incongruent).

To this end, a variation of a previously employed experimental design was adopted (Abraham et al., 2008) where the participants are introduced to one-sentence scenarios in which the mental state verb indicates whether the scenario refers to the beliefs or the desires of a protagonist concerning a given event (Figure 1). A question statement follows the scenario in which the actual reality of that event is introduced which may fulfill or fail to fulfill the protagonist's desire, or may be false or true with reference to the protagonist's belief. Such a design renders it possible to uncover the brain correlates of belief attribution vs. desire attribution, and the processing of incongruent mental states (false belief or unfulfilled desires) relative to congruent mental states (true belief or fulfilled desires).

MATERIALS AND METHODS

Subjects

After excluding two participants due to severe imaging movement artifacts, the sample included 22 right-handed healthy volunteers (11 female; mean age: 26.14; age range: 21–35) with normal or corrected-to-normal visual acuity. All were native German speakers with no history of neurological or psychiatric illness, and none were taking medication at the time of measurement. The participants gave informed consent before participation and the experimental standards were approved by the local ethics committee of the University of Leipzig. Task instructions and a five-minute practice session were given to the participants prior to the fMRI session.

Experimental design

A 2×2 factorial design was employed with the type of mental state to be processed as factor 1 (beliefs, desires) and the congruency or incongruency of the mental state of the protagonist with reference to reality as factor 2 (false/unfulfilled, true/fulfilled). Examples of each of the four resulting conditions are given in Figure 1. The experimental trials (22 per condition) and resting baseline trials (14 trials) were presented in a random trial design. Two further conditions were tested (conditional reasoning, syllogistic reasoning). As the research questions associated with them are beyond the scope of this paper,
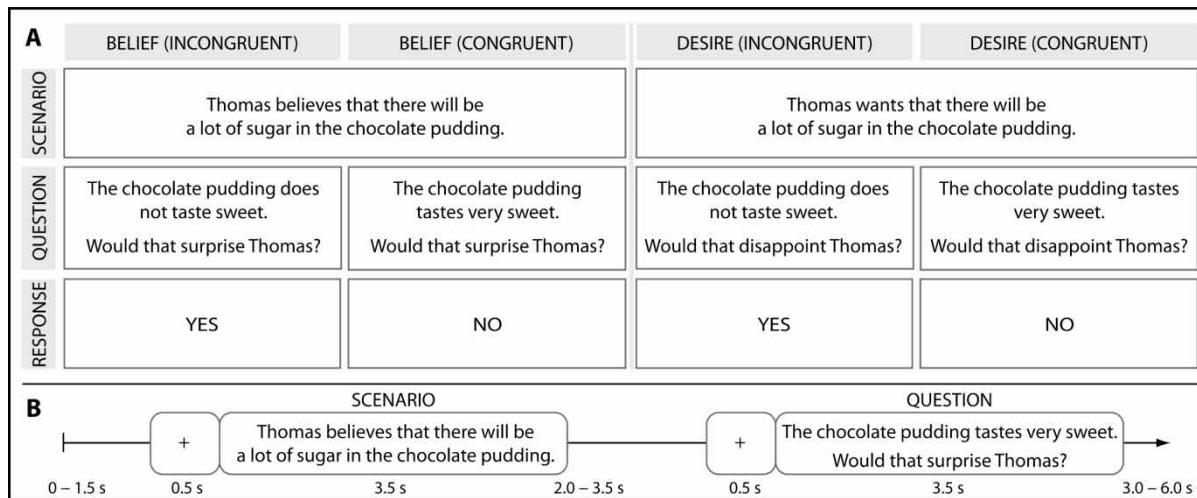
Figure 1. The top rows present examples of scenarios and questions for all experimental conditions. The similarity of the sentences across conditions in the figure is only to serve as an aid in understanding the differences between the conditions. All trials in the experiment were unique in their propositional content. The word-for-word English translations of the original German scenario sentences have been presented to indicate the similarity in the sentence structure between the belief and desire stimuli in German. The alternative question for the belief scenario (Did Thomas expect that?) and the desire scenario (Would that please Thomas?) would lead to opposite responses to those of the belief question and desire question presented. The bottom part of the figure shows a schematic representation of the sequence of events in a trial (trial length: 16 s). To enhance the temporal resolution of the BOLD signal, variable jitter times were inserted before and after each scenario. For the baseline rest condition, a blank screen was presented through the entire trial.

they will not be discussed further except in the context of a non-mental state control condition (syllogistic reasoning) for the fMRI data analysis. With a trial length of 16 s and total of 146 trials, the experimental session spanned 38.93 min.

Stimuli

The stimulus parameters as well as the timing of trial events were determined on the basis of behavioral pilot studies. The experimental material consisted of written sentences in German (font size: 26), which were presented in the centerfield of a screen (resolution: 800×600) spanning three lines (Figure 1). The proper names used in the conditions involving persons were taken from a pool of 20 male and 20 female common German one- and two-syllable names.

The statements for the belief and desire conditions were propositional attitude statements of the form "She believes/desires that p." The mental state verbs that were employed in third person singular form to reflect belief propositional attitudes and desire propositional attitudes were "believes" (*glaubt*) and "wants" (*will*) respectively. Examples of each trial type are given in Figure 1. The content of the "that" clause was

different in each trial. The structural complexity of the belief and desire statements were identical. In fact, the only difference between the belief and desire scenarios was the mental state verb that was employed in the sentence (believes or wants).

During the question phase subjects were presented with one of two types of questions relating to beliefs: "Would that surprise X?" (*Wird X das überraschen?*) and "Is that what X expected?" (*Wird X das erwarten?*). In the same way there were two types of questions related to desires: "Would that please X?" (*Wird das X gefallen?*) and "Would that disappoint X?" (*Wird das X enttäuschen?*). These questions resulted in a yes-or-no response depending on the question type after the altered or unaltered question scenario.

In the syllogistic reasoning (non-mental state) condition, the statements during the scenario phase were in the form "All Xs are Y." An example scenario statement for this condition is "All students at the dance academy own more than three pairs of dance shoes." This was followed by the question phase in which another statement related to the first statement is presented (e.g., Sonja studies dance at the academy) alongside a question (e.g., Does Sonja own only one pair of dance shoes?) to which the participants were required to respond yes or no.

Experimental task

Across all experimental conditions, each trial (Figure 1) began with a fixation cross which was followed by the presentation of single sentence for 3500 ms that introduced a scenario concerning the belief/desire of a protagonist X with reference to a particular event. This scenario phase was followed after a variable delay by the question phase which determine which of the belief trials ($n = 44$) would be allotted to either the false belief ($n = 22$) or the true belief ($n = 22$) conditions, and which of the desire trials ($n = 44$) would be allotted to either the unfulfilled desire ($n = 22$) or the fulfilled desire ($n = 22$) conditions. This question phase included information that rendered the belief/desire introduced in the scenario to be either correct (true/fulfilled) or incorrect (false/unfulfilled), as well as a question regarding the nature of this change, to which the participant was required to respond (Figure 1). This remained on the computer screen for 3500 ms and the participant's task was to respond (yes or no) by pressing the appropriate button (index finger or middle finger) on a response box placed under the right hand. Variable jitter times were inserted before the scenario phase (Jitter 1: 500–2000 ms) and before the question phase (Jitter 2: 2500–4000 ms) to enhance the temporal resolution of the blood oxygenation level-dependent (BOLD) element. We investigated which regions in the brain were engaged during the reading of the scenario (beliefs-scenarios vs. desires-scenarios) as well as during the response phase (beliefs-question vs. desires-question). We also examined the brain regions involved when processing mental state information that, relative to reality, is either incongruent (false-belief-question and unfulfilled-desire-question) or congruent (true-belief-question and fulfilled-desire-question).

MRI scanning procedure

Imaging was carried out on a 3 T Bruker (Ettlingen, Germany) Medspec 30/100 system equipped with the standard birdcage head coil. Participants were placed on the scanner bed in a supine position with their right index and middle fingers positioned on the appropriate response buttons of a response box. The participants' hands were carefully stabilised and form-fitting cushions were utilized to prevent head, arm and hand

movements. They were also provided with ear plugs so that scanner noise would be attenuated. The sentences were presented using the VisuaStim Digital MRI Video System (Resonance Technology, Northridge, USA), which is a high-resolution visor (800×600) comprising two small TFT-screens placed close to the subjects' eyes.

Twenty-four axial slices (19.2 cm field of view; 64×64 pixel matrix; 4 mm thickness; 1 mm spacing; in-plane resolution of 3×3 mm) parallel to the bicommissural line (AC-PC) covering the whole brain were acquired using a single-shot gradient echo-planar imaging (EPI) sequence (TR = 2000 ms; TE = 30 ms; flip angle = 90°; acquisition bandwidth = 100 kHz) sensitive to BOLD contrast. Prior to the functional imaging, 24 anatomical T1-weighted MDEFT images (Norris, 2000; Ugurbil et al., 1993) (data matrix = 256×256; TR = 1300 ms; TI = 650 ms; TE = 10 ms) and 24 T1-weighted EPI images with the same spatial orientation as the functional data were acquired.

fMRI data analysis

The fMRI data were processed using the software package LIPSIA (Lohmann et al., 2001), which contains tools for preprocessing, registration, statistical evaluation and presentation of fMRI data. Functional data were first motion-corrected using a matching metric based on linear correlation. To correct for the temporal offset between the slices acquired in one image, a cubic-spline interpolation was employed. Low-frequency signal changes and baseline drifts were removed using a temporal highpass filter with a cut-off frequency of 1/200 Hz. Spatial smoothing was performed with a Gaussian filter of 5.65 mm full width half maximum (FWHM).

To align the functional data slices onto a three-dimensional stereotactic coordinate reference system, a rigid linear registration was performed with six degrees of freedom (three rotational, three translational). The rotational and translational parameters were acquired on the basis of the MDEFT and EPI-T1 slices to achieve an optimal match between these slices and the individual three-dimensional reference data set. This high-resolution three-dimensional reference data set was acquired for each subject during a previous scanning session. The MDEFT volume data set with 160 slices and 1 mm slice thickness was standardized to the Talairach stereotactic

space (Talairach & Tournoux, 1988). These rotational and translational parameters were subsequently normalized, i.e., transformed by linear scaling to a standard size. The normalized parameters were then used to transform the functional slices using trilinear interpolation so that the resulting functional slices were aligned with the stereotactic coordinate system, thus generating output data with a spatial resolution of $3 \times 3 \times 3$ mm (27 mm^3).

The statistical evaluation was based on a least-squares estimation using the general linear model for serially autocorrelated observations (Friston et al., 1995; Worsley & Friston, 1995). The design matrix was generated with a box-car function, convolved with the hemodynamic response function. Brain activations were analyzed in an event-related design, time-locked to the presentation of the first sentence (for the scenarios) and the second sentence (for the questions) of all presented trials. The six regressors of interest included the belief scenario event, the desire scenario event, the false-belief question event, the true-belief question event, the unfulfilled-desire question event and the fulfilled-desire question event. Trial-by-trial reaction times were included as parameters for the question events of each condition. The resting baseline trials and scenario and question trials of two other unexplored experimental conditions were modelled as separate regressors of no interest. The model equation, including the observation data, the design matrix, and the error term, was convolved with a Gaussian kernel of dispersion of 4 s FWHM to account for the temporal autocorrelation (Worsley & Friston, 1995). In the following, contrast images, i.e., beta value estimates of the raw-score differences between specified conditions were generated for each participant. As all individual functional data sets were aligned to the same stereotactic reference space, the single-subject contrast images were entered into a second-level random-effects analysis for each of the contrasts. One-sample t tests were employed for the group analyses across the contrast images of all subjects which indicated whether observed differences between conditions were significantly distinct from zero. t values were subsequently transformed into Z scores. The results were corrected for multiple comparisons using double thresholding, i.e., a combination of single voxel probability thresholding on one hand (here $z = 2.33$, $p < .01$), and cluster-size and cluster-z-value thresholding on the other (here $p < .05$), which is computed using Monte-Carlo simulations (Lohmann et al., 2008).

Inclusive mask analyses were carried out from the corrected one-sample t-tests. In each inclusive mask analysis, the statistic parametric map of the random-effects analysis of the experimental condition A vs. control condition direct contrast was used as an inclusive mask in the random-effects analysis of the experimental condition A vs. experimental condition B direct contrast. The findings that result from an inclusive masked analysis indicate which brain areas were significantly activated for experimental condition A relative to experimental condition B, but only if the same regions were also more highly activated in experimental condition A relative to the control condition C.

For percentage signal change (PSC) analyses of the BOLD response, the mean PSC over the event (after the first fixation cross for the "scenario" event and the second fixation cross for the "question" event) was extracted from the peak voxel within significantly activated brain areas for the experimental and resting baseline conditions. The mean PSC of a voxel for each condition was calculated in relation to the mean signal intensity of that voxel across all time steps.

The list of contrasts included (a) belief-scenario > desire-scenario (inclusive mask: belief-scenario > control-scenario); (b) desire-scenario > belief-scenario (inclusive mask: desire-scenario > control-scenario); (c) belief-question > desire-question (inclusive mask: belief-question > control-question); (d) desire-question > belief-question (inclusive mask: desire-question > control-question); (e) incongruent > congruent mental states; (f) false-beliefs > true beliefs; (g) incompatible desires > compatible desires; and (h) conjunction analysis (belief-question > control-question) + (desire-question > control-question).

RESULTS

Behavioral results

The descriptive data (mean and standard deviation) for reaction time (RT) and percentage correct responses (PCR) across all experimental conditions are presented in Table 1. A 2×2 repeated measures ANOVA on PCR revealed only a significant main effect for mental state type, $F_{1,21} = 23.25$, $p < .001$, indicating that the participants were less accurate when answering

TABLE 1

Descriptive data (mean and standard deviation) of the behavioral measures across all conditions: False Beliefs, True Beliefs, Unfulfilled Desires, Fulfilled Desires and the Non-Mental State Control condition

	RT (ms)		PCR (%)	
	Mean	SD	Mean	SD
Belief (False)	2633.14	376.03	89.88	8.05
Belief (True)	2668.09	353.45	91.12	6.94
Desires (Unfulfilled)	2595.09	410.19	94.63	4.78
Desires (Fulfilled)	2523.73	367.27	96.90	4.52
Control	2912.23	386.27	91.73	5.89

Notes: RT: reaction time; PCR: percentage of correct responses.

questions related to belief mental states (false and true) compared to desire mental states (unfulfilled and fulfilled). The main effect for mental state congruency ($F_{1,21} = 1.69$, $p = .21$) as well as the interaction effect between mental state type and congruency ($F_{1,21} = 0.14$, $p = .71$) were non-significant. A 2×2 repeated measures ANOVA on RT also showed a significant main effect for mental state type on correct trials, $F_{1,21} = 23.01$, $p < .001$, revealing that the participants took longer to respond to questions related to belief mental states (false and true) compared to desire mental states (unfulfilled and fulfilled). While the main effect for mental state congruency ($F_{1,21} = 0.61$, $p = .44$) was non-significant, a significant interaction effect between mental state type and congruency was found on the RT measure, $F_{1,21} = 9.38$, $p = .006$. Analyses using *t*-tests revealed that this interaction effect was driven by longer RTs that were associated with the unfulfilled desire condition compared to the fulfilled desire condition ($t_{21} = 2.08$, $p = .05$).[3]

In summary, the behavioral findings demonstrate that the processing of belief representations was more cognitively demanding (lengthier response times and lower response accuracy) than the processing of desire representations. It should be kept in mind that these significant differences reflected consistent but small differences in the dependent variables as the range for response accuracy across all conditions ranged between 89.9% and 96.9% and that of response duration between 2.52 and 2.69 s.

In order to rule out the possibility that the differences between brain regions engaged during

___

[3] As this was an unexpected finding and only approaching significance, we do not discuss this result further.

the different types of mental state reasoning could be attributable simply to RT differences, the contrasts of the belief and desire conditions vs. the non-mental state syllogistic reasoning condition were used as inclusive masks for all the fMRI analyses (see below for details). This non-mental state condition was an ideal control condition as it was associated with significantly longer RTs in comparison to the false belief ($t_{21} = 9.04$, $p < .0001$), true belief ($t_{21} = 7.81$, $p < .0001$), incompatible desire ($t_{21} = 8.79$, $p < .0001$) and compatible desire ($t_{21} = 12.32$, $p < .0001$) conditions.

fMRI results (mental state reasoning)

A conjunction analysis was carried out between the whole brain corrected direct contrasts of belief-question > control-question and desire-question > control-question (Table 2). Brain areas generally activated for mental state reasoning in comparison to nonmental state reasoning included the medial prefrontal and anterior cingulate cortices as well as the posterior cingulate cortex and precuneus (Figure 2).

fMRI results (belief)

Two inclusive masked contrasts were carried out to verify which regions of the brain are involved when processing belief ascriptions in comparison to desire ascriptions (Table 3, Figure 3). One contrast demonstrated the brain areas that were involved in belief processing during the scenario phase (belief-scenario vs. desire-scenario) and the other contrast demonstrated the brain areas that were involved in belief processing during the question phase (belief-question vs. desire-question). Only lateral frontal lobe regions were found to be involved during belief processing in both the scenario phase and the question phase (Table 3). Although the activations extended bilaterally, they were more extensive in the right hemisphere. The engaged areas included the anterior prefrontal regions (BA 10), inferior prefrontal regions (BA 45/47) and middle prefrontal regions (BA 9/46). The general pattern of activated regions during belief processing was very similar across both the scenario phase and the question phase, but with a slightly more extensive engagement of frontal regions in the question phase.

TABLE 2

Mental state reasoning: Anatomical specification, Talairach coordinates, maximum Z value and volume (mm^3) of the significantly activated areas in the conjunction analysis of the Belief-Question $>$ Control-Question contrast and the Desire-Question $>$ Control-Question contrast

Brain areas	L/R	x	y	z	mm^3	Z
Medial PFC/Anterior cingulate (BA 32/24)	L	−5	7	36	15795	4.72
Medial PFC/ACC (BA 32/10)	R	7	34	12	lm	4.41
Superior/Middle frontal g (BA 9)	L	−23	40	24	lm	2.63
Superior/Middle frontal g (BA 9)	R	22	40	24	lm	3.19
Posterior cingulate/Precuneus	R	7	−32	42	4995	3.60
Inferior parietal l/Supramarginal g	R	58	−38	21	28296	4.69
Inferior parietal l	R	55	−29	36	lm	4.67
Inferior frontal g (BA 45)	R	43	28	9	lm	3.15
Insula	R	43	−2	9	lm	4.31
Insula	L	−47	−5	9	5454	3.86
Insula	L	−35	1	15	lm	3.67
Supramarginal g	L	−50	−32	24	3699	3.43
Caudate nucleus	L	−17	10	−3	1485	3.55
Caudate nucleus	R	13	16	0	1404	3.45

Notes: L: left hemisphere; R: right hemisphere; lm: local maxima; g: gyrus; l: lobule. The results were corrected for multiple comparisons at the cluster level ($p < .05$).

fMRI results (desire)

Just as in the case of belief processing, two inclusive masked contrasts were carried out to verify which regions of the brain are involved when processing desire ascriptions in comparison to belief ascriptions (Table 4, Figure 4). Here again, one contrast demonstrated the brain areas that were involved in desire processing during the scenario phase (desire-scenario vs. belief-scenario) and the other contrast demonstrated the brain areas that were involved in desire processing during the question phase (desire-question vs. belief-question). Unlike in the case of belief

processing, desire processing during the question phase was associated with a much larger extent of activations in different regions of the brain compared to the desire processing during the question phase (Table 4). For instance, a more extensive (and bilateral) expanse of regions along lateral middle and inferior temporal regions were also more activated during the desire-question phase compared to the desire-scenario phase. Regions that were only activated during the desire-question phase and not the desire-scenario phase included the medial prefrontal cortex, anterior cingulate cortex, hippocampal formation and insula.
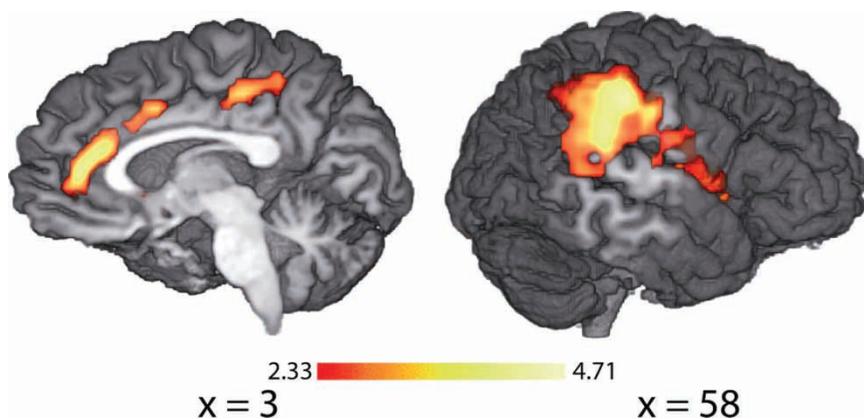

Figure 2. Processing of mental states: The activations resulted from the conjunction analysis of Belief-question $>$ Control-question contrast and the Desire-question $>$ Control-question contrast.

TABLE 3

Belief mental state processing: Anatomical specification, Talairach coordinates, maximum Z value and volume (mm^3) of the significantly activated areas in the Belief-Scenario > Desire-Scenario contrast (inclusive mask: Belief-Scenario > Control-Scenario contrast) and the Belief-Question > Desire-Question contrast (inclusive mask: Belief-Question > Control-Question contrast)

Brain area	L/R	X	y	z	mm^3	Z
Belief Scenario > Desire Scenario						
Middle/inf frontal g (BA 10/45/47)	L	−44	31	21	1188	3.57
Superior/middle frontal g (BA10)	L	−26	58	21	216	3.72
Superior/middle frontal g (BA 10)	R	22	55	24	1053	4.11
Superior/middle frontal g (BA 9)	R	22	46	33	108	3.00
Belief Question > Desire Question						
Inferior/Middle frontal g (BA 10/47)	R	37	43	9	2187	5.01
Middle frontal g (BA 10)	R	25	52	21	lm	4.61
Middle/Superior frontal g (BA 8/9)	R	34	40	33	lm	3.59
Middle/Inferior frontal g (BA 46/45)	L	−44	31	24	405	3.50
Middle/Superior frontal g (BA 8/9)	L	−29	43	36	108	2.80

Notes: L: left hemisphere; R: right hemisphere; lm: local maxima; g: gyrus. The results were corrected for multiple comparisons at the cluster level ($p < .05$).

fMRI results (incongruent vs. congruent)

To find out which brain regions were more strongly engaged during incongruent intentional state processing, a direct contrast was carried out in which the incongruent conditions (false-beliefs and un-fulfilled-desires) were contrasted with congruent conditions (true-beliefs and fulfilled-desires). This contrast is naturally only of the question phase because (a) it is not until the question phase that the congruency of the beliefs/desires with reality is known, and (b) the belief scenario phase is identical for false and true beliefs just as the desire scenario phase is identical for unfulfilled and fulfilled desires. Only five regions were found to be activated in this contrast, all of which were situated along the medial wall of the brain (Table 5,
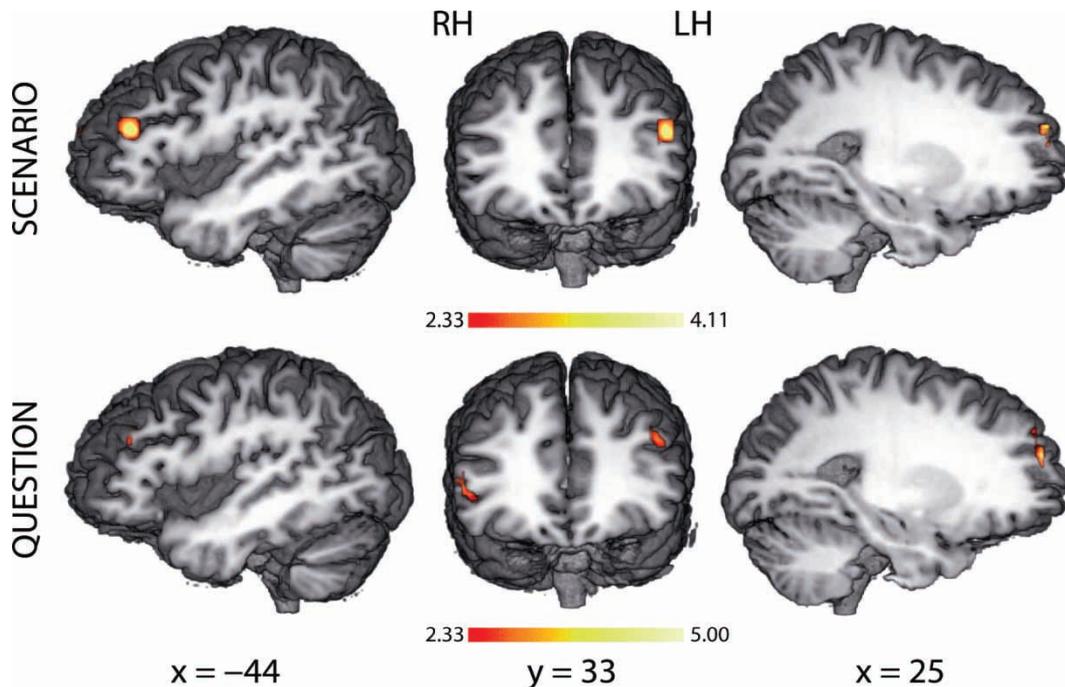

Figure 3. Processing of beliefs: The activations in the top panel are those resulting from the Belief-scenario > Desire-scenario contrast; the activations in the bottom panel are those resulting from the Belief-question > Desire-question contrast. The results were corrected for multiple comparisons at the cluster level ($p < .05$).

TABLE 4

Desire mental state processing: Anatomical specification, Talairach coordinates, maximum Z value and volume (mm^3) of the significantly activated areas in the Desire-Scenario >Belief-Scenario contrast (inclusive mask: Desire-Scenario >Control-Scenario contrast) and the Desire-Question >Belief-Question contrast (inclusive mask: Desire-Question >Control-Question contrast)

Brain area	L/R	x	y	z	mm^3	Z
Desire Scenario >Belief Scenario						
Inferior temporal g	R	49	−71	9	1053	3.42
Inferior temporal g	R	49	−65	−3	lm	3.10
Inferior temporal/Occipital g	R	43	−77	−3	81	3.62
Occipital g	R	40	−80	12	432	3.27
Middle temporal g	L	−35	−53	18	189	2.82
Inferior parietal lobe	L	−32	−68	24	81	2.88
Superior temporal g	L	−44	−14	−9	81	2.54
Desire Question >Belief Question						
Medial PFC/Anterior Cingulate (BA 10/9/32)	L	−2	34	0	4590	3.81
Middle/inferior temporal g (BA 21/20)	R	40	−11	−12	2403	4.72
Middle/superior temporal g	R	43	−50	18	108	2.92
Superior temporal g	L	−44	−17	−3	1566	3.51
Insula	R	49	−8	0	lm	3.41
Superior temporal g	L	−47	−47	24	108	2.83
Superior temporal g	L	−56	−29	12	135	2.73
Middle temporal/occipital g (BA 39/40/19)	R	49	−71	12	702	4.33
Insula	L	−32	10	−12	189	2.76
Hippocampal formation/Amygdala	R	25	−5	−15	162	3.05
Hippocampal formation	R	25	−20	−12	243	3.08
Inferior temporal g	R	37	−17	−33	lm	3.87
Postcentral g/Precentral g	R	34	−26	54	2079	3.49
Superior parietal lobule	R	25	−35	57	lm	3.44
Occipital g	R	10	−95	27	648	3.43
Occipital g	L	−6	−98	18	81	3.02
Superior temporal g	L	−50	−32	15	135	59
Subthalamic nucleus	L	−5	−11	−3	297	287

Notes: L: left hemisphere; R: right hemisphere; lm: local maxima; g: gyrus). The results were corrected for multiple comparisons at the cluster level ($p < .05$).

Figure 5). The three activated peaks in the fronto-median wall comprised the pregenual medial PFC (BA 10/32) bilaterally and the dorsal medial PFC (BA 9/10) unilaterally, whereas the activated peaks along the posteromedian wall was situated in the posterior cingulate and retrosplenial regions bilaterally (BA 31/23). PSC analyses were carried out to determine the pattern of activation in these regions. As shown in Figure 5, unlike the other regions that were only modulated differentially by the main effect factors of type of mental state and type congruency of mental state, only the activation in the dorsal medial prefrontal cortex showed a significant interaction effect, $F_{1,19} = 17.8$, $p < .001$. This indicated that this region was most strongly engaged during the false belief condition for which the activations were significantly higher than in the true belief condition, $t_{19} = 3.5$, $p = .002$, the unfulfilled desire condition, $t_{19} = 3.09$, $p = .006$, and the fulfilled desire condition, $t_{19} = 2.33$, $p = .031$. Direct contrasts were also carried out within intentional states (Table 5). The processing of false

beliefs in comparison to true beliefs resulted in activations in the abovementioned areas among others, whereas no region was significantly activated (i.e., survived the correction threshold) when contrasting the processing of unfulfilled-desires with that of fulfilled-desires.

DISCUSSION

The objectives of the present study were twofold. Using one experimental paradigm, we aimed to determine which brain regions were differentially involved when engaging reasoning about belief vs. desire intentional states. In addition, we sought to uncover which regions were preferentially engaged when processing contexts containing incongruent mental state information (false beliefs, unfulfilled desires) vs. congruent mental state information (true beliefs, fulfilled desires). The question was open as to whether one or more of the commonly implicated brain regions in mental
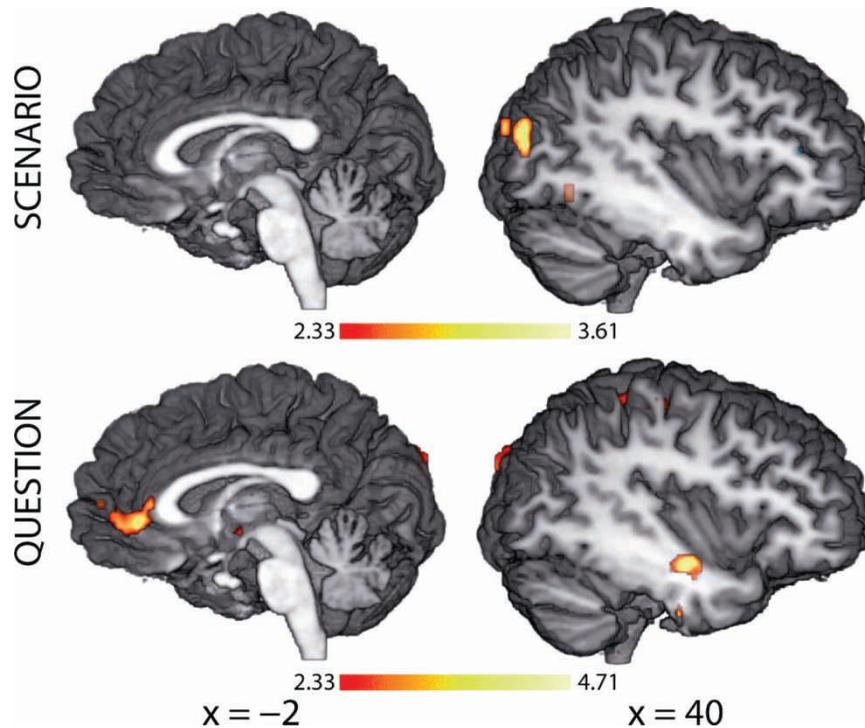
Figure 4. Processing of desires: The activations in the top panel are those resulting from the Desire-scenario >Belief-scenario contrast, whereas the activations in the bottom panel are those resulting from the Desire-question >Belief-question contrast. The results were corrected for multiple comparisons at the cluster level ($p < .05$).

state reasoning would be additionally selective for one or the other type of intentional state and/ or whether additional brain regions would be recruited.

Beliefs vs. desires

Comparisons of both belief and desire mental reasoning compared to non-mental state syllogistic reasoning revealed the engagement of brain regions commonly found to be involved in mentalizing or theory of mind such as the medial prefrontal and anterior cingulate cortices as well as the posterior cingulate cortex and the precuneus alongside the inferior parietal and temporal lobe regions. This fits with the wealth of fMRI literature on mental state reasoning (Frith & Frith, 2003, 2006; Gallagher & Frith, 2003; Saxe, 2006; Saxe et al., 2004a).

The first objective of the current study was to determine how our understanding of distinct mental states such as beliefs and desires is represented in the brain. The experimental paradigm allowed us to determine which brain areas are involved when processing information concerning a protagonist's beliefs and desires with

reference to an event (scenario phase) as well as which brain areas are involved when using this information to predict what the protagonist would believe or desire based on whether the reality is in line with the belief/desire or not (question phase).

The behavioral findings, which only pertained to the question phase, revealed that the processing of beliefs was associated with greater demands on information processing than the processing of desires, as evidenced by longer response times and lower performance accuracy associated with the former. These significant findings, although small, are striking especially given that the stimuli for the belief and desire conditions were identical in terms of linguistic complexity and that they did not vary in any other dimension except in that they tapped either beliefs or desires. In the scenario phase, a statement was defined as belonging to the belief or desire category based on the presence of one of two third person singular verbs ("believes" or "wants"). In the question phase, the reality information was the same for both beliefs and desires and the only difference was the nature of the questions posed as indicated by the adjective (belief: expected/surprised, desire: pleased/disappointed). Given these rather minor differences in

TABLE 5

Incongruent mental state processing: Anatomical specification, Talairach coordinates, maximum Z value and volume (mm^3) of the significantly activated areas in the Incongruent Beliefs & Desires > Congruent Beliefs & Desires contrast, False-beliefs > True-beliefs contrast, and the Unfulfilled-desires > Fulfilled-desires contrasts

Brain area	L/R	x	y	z	mm^3	Z
Incongruent > Congruent						
Ventral medial PFC (BA 10/32)	L	−11	49	15	3429	4.14
Ventral medial PFC (BA 10/32)	R	7	49	6	lm	3.05
Dorsal medial PFC (BA 10/9)	L	−8	61	27	lm	3.08
Posterior cingulate (BA 31/23)	L	−2	−65	21	4590	4.00
Posterior cingulate (BA 31/23)	L	−17	−62	18	lm	3.83
False Beliefs > True Beliefs						
Medial PFC (BA 9/32)	L+R	−2	58	21	5940	4.05
Ventral medial PFC (BA 10/32)	L+R	−5	46	3	lm	3.60
Posterior cingulate (BA 31/23)	L+R	−2	−65	21	6588	3.68
Posterior cingulate (BA 31/23)	L	−17	−47	15	lm	3.92
Posterior cingulate & retrosplenial cortex	R	19	−50	12	lm	3.03
Posterior cingulate & retrosplenial cortex	L	−17	−47	9	lm	3.41
Hippocampal formation & amygdala	R	28	−23	−9	7317	4.56
Hippocampal formation & amygdala	R	19	−2	−9	lm	3.66
Hippocampal formation & amygdala	R	28	−23	−9	lm	4.56
Hippocampal formation	L	−29	−17	−6	lm	4.90
Insula	L	−47	−11	21	11178	4.23
Superior temporal gyrus	L	−50	−8	0	lm	3.17
Putamen	L	−20	10	−3	lm	3.83
Premotor cortex	L	−23	−32	69	3294	3.85
Precentral gyrus	L	−29	−20	57	lm	3.24
Precentral gyrus	L	−47	−35	60	lm	3.10
Inferior frontal gyrus (BA 45)	R	43	25	12	918	4.57
Unfulfilled Desires > Fulfilled Desires (none)						

Notes: L: left hemisphere; R: right hemisphere; lm: local maxima. The results were corrected for multiple comparisons at the cluster level ($p < .05$).

stimuli, the behavioral differences between processing beliefs and desires that resulted are noteworthy because they suggest that there is a fundamental difference in the means by which we deal with information pertaining to beliefs and desires in that a greater information processing load is tied together with belief reasoning.

The fMRI evidence in fact revealed that the brain regions that were activated when processing belief state information during the question phase (when an intentional ascription was made) were also manifest during the scenario phase[4] (when intentional state information was read and kept in mind). This suggests that there is an engagement of common brain regions prior to (perhaps in anticipation of further task requirements) and

during the making of belief intentional ascriptions. Behavioral findings do in fact indicate that theory of mind inferences are relatively non-automatic as they are reported less promptly when asked incidentally after hearing stories rather than when explicitly informed beforehand (Apperly et al, 2006a).

Brain regions involved in belief processing relative to desire processing were confined to the lateral aspect of frontal lobe with the engagement of anterior prefrontal regions, middle frontal regions and inferior frontal regions. These lateral prefrontal regions are known to be involved in exerting executive or cognitive control, which refers to the ability to synchronize thoughts and actions in relation to internal goals at all stages of complexity from rudimentary to abstract levels (Koechlin, Ody, & Kouneiher, 2003; Koechlin & Summerfield, 2007). These executive aspects of control are necessary for diverse cognitive processes such as working memory and reasoning (Christoff et al., 2001; Gilbert & Burgess, 2008; Ruff, Knauff, Fangmeier, & Spreer,

---

[4] Events pertaining to the scenario stage (where mental state information is being conveyed) and the question stage (where explicit mental state reasoning on part of the participant is required) were analyzed separately because the similarities and differences between the recruitment of brain regions as a function of which stage of mental state reasoning one is engaged in have not been systematically explored so far.
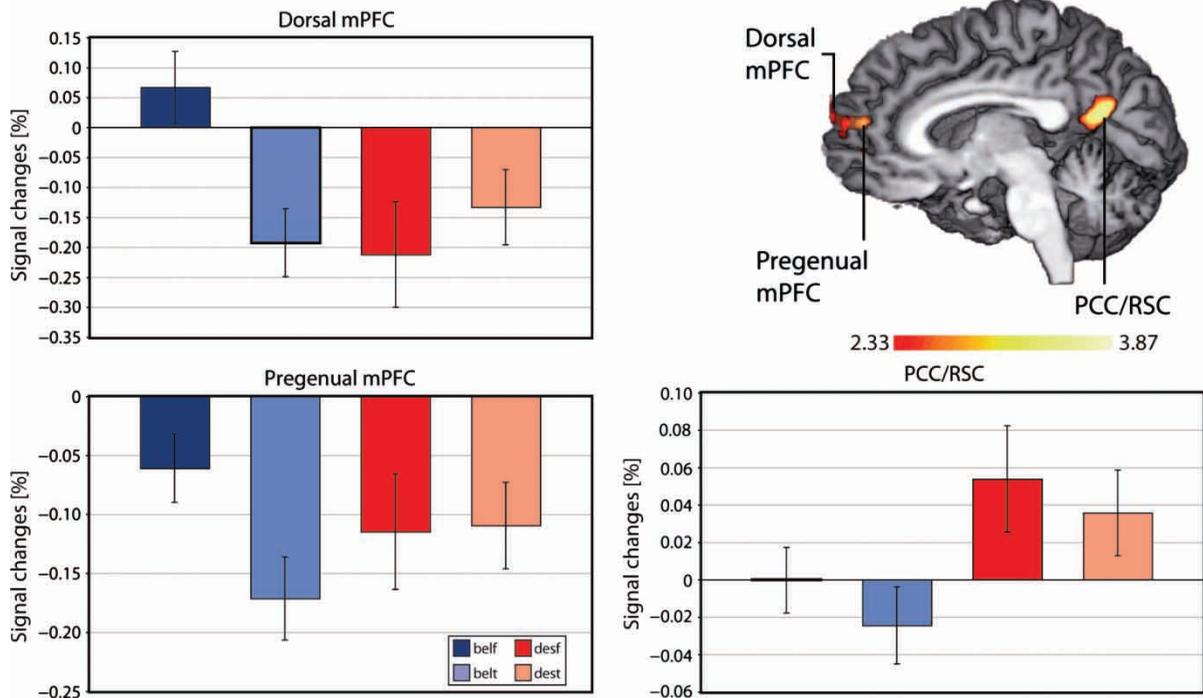
Figure 5. Processing of incongruent mental states: The activations are those that resulted from the contrast of incongruent mental states (false belief and unfulfilled desires) with congruent mental states (true belief and fulfilled desires). The results were corrected for multiple comparisons at the cluster level ($p < .05$). Percent signal change plots are presented for regions in the dorsal medial prefrontal cortex (dmPFC: $-8, 61, 27$), pregenual mPFC ($-2, 58, 21$), and posterior cingulate/retrosplenial cortex (PCC: $-2, -65, 21$), as labeled. The zero point represents the resting baseline.

2003) as well as in social settings (Weissman, Perkins, & Woldorff, 2007). The lateral PFC has been therefore proposed to underlie the monitoring and manipulation of information, which is stored in posterior association areas, in service of a goal (Muller & Knight, 2006; Petrides, 2005). Most studies on cognitive control, though, demonstrate a stronger involvement of the left hemisphere, whereas more extensive right hemisphere activation was found in the current study. This might be due to the possibility that the information to be manipulated and monitored in the case of belief representations is more abstract compared to commonly studied cognitive control tasks. More dominant right hemisphere frontal activation has in fact been reported in studies in response to processing more abstract information (Aziz-Zadeh, Kaplan, & Iacoboni, 2009; Stringaris et al., 2006). This idea also fits with the finding of no significant activations in other regions that often accompany lateral frontal involvement, such as the posterior parietal cortex, a region which is involved when orienting towards task-relevant spatial and sensory information in the environment.

Desire processing, in comparison to belief processing, did not show as extensive overlapping regions during the scenario and question phase. Activations in the lateral inferior temporal and the inferior parietal cortices in the scenario phase were supplemented by the recruitment of a network of other brain regions including the hippocampal formation, insula, medial prefrontal and anterior cingulate regions during the question phase. The latter regions include parts of the brain that are known to be typically engaged during autobiographical episodic memory retrieval as well as self-referential thinking (Cavanna & Trimble, 2006; Gilbert et al., 2006; Maguire, 2001; Svoboda, McKinnon, & Levine, 2006; Vogt & Laureys, 2005), and have also been widely implicated in neuroimaging studies on mental state reasoning.[5] Some of the regions involved in desire processing

---

[5] There were also additional activations during desire relative to belief reasoning in occipital and fusiform regions. These unexpected results could reflect condition specific lower-level phenomena, such as different eye movement trajectories as a function of different reading strategies, or higher-level phenomena such as greater visual imagery. Further experiments and behavioral indices will be required to verify which of these alternatives is more plausible.

compared to belief processing also form part of the "default network" of the brain, which refers to a group of brain regions that are customarily more engaged during passive periods within experiments, such as at rest or when performing cognitively undemanding tasks compared to highly demanding tasks (e.g. Buckner, Andrews-Hanna & Schacter, 2008; Raichle et al., 2001). It must be noted though that parts of what constitutes the "core" default network (Buckner et al., 2008), such as the posterior cingulate/retrosplenial cortices, were not found to be significantly more engaged during desire processing.

The findings therefore reveal that there are dissociations between the manner in which the processing of beliefs and desires is represented in the brain, with belief processing drawing on conceptual and relational processing related brain structures whereas desire processing is associated with the recruitment of structures involved in self-referential processing.

It may be argued that the desire vs. belief ascription differences we report could be a result of the behavioral differences between the conditions. While this issue does impose some limitations in the interpretations of the fMRI results, an argument against it is that the pattern of activations was consistent during the scenario phase (when no behavioral response was required) and the question phase (when a behavioral response was required), especially in the case of beliefs. More importantly though, inclusive masked analyses were employed by using contrasts against a non-mental syllogistic condition as an inclusive mask when contrasting beliefs and desires. As this non-mental condition was associated with the largest RTs, it cannot be argued that the activation differences seen between beliefs and desires are simply a function of RT differences.

The behavioral differences between the conditions are interesting in themselves, especially within the context of the developmental asymmetry involved in the understanding of beliefs and desires. It is important to understand what the ramifications of this effect are and how it relates to other aspects of cognition. One possible phenomenon it could relate to is the widely studied "self-reference effect" (SRE) in social cognition. SRE refers to the tendency that information is better remembered when it relates to oneself in comparison to information that involves less personal relevance (for a review, see Symons & Johnson, 1997). It could be the case that within our paradigm, desires were more effortlessly processed as

they were not counterintuitive and could readily be aligned to one's own preferences that were likely to be true of that situation. One way to test this hypothesis would be to assess desires in scenarios that are either incompatible or compatible with one's own preferences in a given situation.

False beliefs/unfulfilled desires vs. true beliefs/fulfilled desires

The second objective of the present study was to characterize which regions are involved when having to integrate mental state information that is incongruent with reference to reality. Incongruent mental states in our paradigm were those in which a protagonist's beliefs or desires were incongruent with that of reality (false beliefs, unfulfilled desires) and these were compared to congruent mental states, where a protagonist's beliefs or desires are true or congruent with reality (true beliefs, fulfilled desires).

No previous neuroimaging study has contrasted false and true beliefs as well as unfulfilled and fulfilled desires within the same experimental paradigm. While pregenual medial prefrontal regions were more strongly modulated by the processing of both incongruent mental states, only a dorsal medial prefrontal region (BA 9/10) showed a selective pattern such that it was most strongly activated during the false belief condition. This region is commonly reported to be involved in studies on mental state reasoning, which customarily tap false/true belief reasoning relative to non-theory-of-mind control tasks (Frith & Frith, 1999; Gallagher & Frith, 2003). The activated area is located in the vicinity of the region commonly referred to as the anterior paracingulate cortex, which has been postulated to be the critical region that underlies the process of decoupling during mental state reasoning (Gallagher & Frith, 2003). Decoupling refers to the process whereby the representation of the belief, which may or may not be contrary to reality, is kept separate from the representation of the reality (Leslie, 1987). The current finding that the dorsal medial prefrontal cortex is most highly activated during false belief ascriptions compared to ascription of true beliefs, unfulfilled desires and fulfilled desires can thus be well aligned with established ideas and previous findings within the literature.

Unlike the case of false beliefs, no brain region was found to be exclusively activated when

processing information relating to unfulfilled desires. This fits well with the idea that understanding desires constitutes early developing concepts of theory of mind that do not necessitate fully developed meta-representational capabilities (Perner, 1993). Crucially, desires do not share the logical normative structure of beliefs that aim at truth: Beliefs ought to be true and consequently get revised in light of evidence that they are false (reflecting their mind-to-world direction of fit). Desires, in contrast, though aiming at fulfillment, do not have a normative default of being fulfilled, and therefore it is not the case that they ought to be revised in light of evidence that they are unfulfilled (on the contrary, the world ought to be changed, so to speak – reflecting the world-to-mind direction of fit of desires).

To date, only one prior study has directly compared understanding false beliefs with true beliefs (Sommer et al., 2007). The findings of this study conflict with those of the present study in that they reported activations in a posterior region along the medial wall in the dorsal anterior cingulate cortex as well as in the right lateral prefrontal cortex (BA 9/10) and the right TPJ in association with false belief reasoning. The differences between the current paradigm and the Sommer et al. (2007) paradigm may explain some of the differences between the findings. First of all, the stimuli in the previous study were designed such that there were only four types of trials, all of which were modelled on the Sally-Anne task. As each of these trials was conducted 20 times, it is likely that the demands on belief attribution reduced considerably over time. A reduced demand on belief attribution may also explain why the results of the conjunction analysis of the true beliefs and false beliefs vs. a non-belief story phase in this study did not result in the engagement of brain regions customarily found in functional imaging studies on theory of mind. In contrast, in the present study, each situation had to be assessed anew because the content of the beliefs differed from trial to trial, thereby keeping the demand on belief attribution constant.

A second reason for the disparity of the findings could be the fact that the Sommer et al. (2007) paradigm used pictorial stimuli depicting a scene, whereas simple first-order propositions were used in the current study. Previous work has shown that the right TPJ, for instance, is particularly sensitive to aspects of mental state reasoning that are tied to representations of persons-in-space as well as greater levels of representational complexity (Abraham et al., 2008). Such findings might explain the lack of right TPJ involvement in the present study as only simple propositional attitude statements were investigated.

A third factor to keep in mind is existing theoretical ideas on reasoning about intentional states and its relation to brain function. One of the most influential regarding the brain bases of mental state reasoning (Saxe, 2006) claims on the basis of neuroimaging (Saxe & Kanwisher, 2003; Saxe & Powell, 2006) and neuropsychological evidence (Samson et al., 2005) that the TPJ region is important for the representation of intentional mental states generally—which include both beliefs and desires (Saxe & Kanwisher, 2003; Saxe & Wexler, 2005) (for an alternative standpoint, see Frith & Frith, 2003, 2006). Furthermore, false belief tasks have largely been contrasted not with true beliefs, but with other control tasks such as false photograph task or physical stories (e.g., Saxe & Kanwisher, 2003). So an undifferentiated response in the TPJ across first-order false and true belief processing as well as unfulfilled and fulfilled desire processing within the current study is not incompatible with prominent ideas in the literature.

Outlook: Beliefs, desires and beyond

The two paradigmatic kinds of propositional attitudes we ascribe in our folk psychology and make use of in rational action explanation are beliefs and desire. The present study, the first to directly compare structurally analogous ascriptions of beliefs and desires, suggests there are different neurological substrates for attributions of these two kinds of attitudes. It should be noted, however, that beliefs and desires are not the only kinds of propositional attitudes that play important roles in our folk psychology. First, there are more complex, mixed kinds of attitudes that share some cognitive element with beliefs (some mind-to-world direction of fit) and some conative element with desires (some world-to-mind direction of fit). Propositional emotions such as ''to be happy that . . .'' fall in this category. When I am happy that my football team won, this involves a cognitive element (my belief that they won) and a conative one (my approval of their winning). Second, there are propositional attitudes that have no obvious direction of fit at all, such as

imagination (if I imagine that p, I neither say anything about the world nor about how I would like it to be). A valuable direction for future research on mental state reasoning would therefore be to contrast the ascription of such and other propositional attitudes and emotions to ascertain the parallels and differences between their neurophysiological bases.

REFERENCES

Abraham, A., Werning, M., Rakcozy, H., Von Cramon, D. Y., & Schubotz, R. I. (2008). Minds, persons, and space: An fMRI investigation into the relational complexity of higher-order intentionality. *Consciousness and Cognition*, *17*, 438–450.

Anscombe, G. E. M. (1957). *Intention*. Oxford, UK: Basil Blackwell.

Apperly, I. A., Riggs, K. J., Simpson, A., Samson, D., & Chiavarino, C. (2006a). Is belief reasoning automatic? *Psychological Science*, *17*(10), 841–844.

Apperly, I. A., Samson, D., Chiavarino, C., Bickerton, W. L., & Humphreys, G. W. (2006b). 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*(2), 300–321.

Aziz-Zadeh, L., Kaplan, J. T., & Iacoboni, M. (2009). ''Aha!'': The neural correlates of verbal insight solutions. *Human Brain Mapping*, *30*(3), 909–916.

Bar, M. (2007). The proactive brain: Using analogies and associations to generate predictions. *Trends in Cognitive Sciences*, *11*(7), 280–289.

Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The brain's default network: Anatomy, function, and relevance to disease. *Annals of the New York Academy of Sciences*, *1124*, 1–38.

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*(3), 314–325.

Cavanna, A. E., & Trimble, M. R. (2006). The precuneus: A review of its functional anatomy and behavioural correlates. *Brain*, *129*(3), 564–583.

Christoff, K., Prabhakaran, V., Dorfman, J., Zhao, Z., Kroger, J. K., Holyoak, K. J., et al. (2001). Rostrolateral prefrontal cortex involvement in relational integration during reasoning. *NeuroImage*, *14*(5), 1136–1149.

Ciaramidaro, A., Adenzato, M., Enrici, I., Erk, S., Pia, L., Bara, B. G., et al. (2007). The intentional network: How the brain reads varieties of intentions. *Neuropsychologia*, *45*(13), 3105–3113.

Davidson, D. (1963). Actions, reasons, and causes. *Journal of Philosophy*, *60*, 685–700.

Decety, J., & Lamm, C. (2007). The role of the right temporoparietal junction in social interaction: How low-level computational processes contribute to meta-cognition. *Neuroscientist*, *13*(6), 580–593.

Dennett, D., & Haugeland, J. (1987). Intentionality. In R. L. Gregory (Ed.), *The Oxford companion to the mind* (pp. 161–164). New York: Oxford University Press.

Ferstl, E. C., & von Cramon, D. Y. (2002). What does the frontomedian cortex contribute to language processing: Coherence or theory of mind? *NeuroImage*, *17*(3), 1599–1612.

Fletcher, P. C., Happé, F., Frith, U., Baker, S. C., Dolan, R. J., Frackowiak, R. S., et al. (1995). Other minds in the brain: A functional imaging study of ''theory of mind'' in story comprehension. *Cognition*, *57*(2), 109–128.

Friston, K., Holmes, A., Worsley, K., Poline, J., Frith, C., & Frackowiak, R. (1995). Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, *2*, 189–210.

Frith, C. D., & Frith, U. (1999). Interacting minds: A biological basis. *Science*, *286*(5445), 1692–1695.

Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron*, *50*(4), 531–534.

Frith, U., & Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *358*(1431), 459–473.

Frith, U., Morton, J., & Leslie, A. M. (1991). The cognitive basis of a biological disorder: Autism. *Trends in Neuroscience*, *14*(10), 433–438.

Gallagher, H. L., & Frith, C. D. (2003). Functional imaging of 'theory of mind'. *Trends in Cognitive Sciences*, *7*(2), 77–83.

Gallagher, H. L., Happé, 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*(1), 11–21.

German, T. P., Niehaus, J. L., Roarty, M. P., Giesbrecht, B., & Miller, M. B. (2004). Neural correlates of detecting pretense: Automatic engagement of the intentional stance under covert conditions. *Journal of Cognitive Neuroscience*, *16*(10), 1805–1817.

Gilbert, S. J., & Burgess, P. W. (2008). Executive function. *Current Biology*, *18*(3), R110–R114.

Gilbert, S. J., Spengler, S., Simons, J. S., Steele, J. D., Lawrie, S. M., Frith, C. D., et al. (2006). Functional specialization within rostral prefrontal cortex (area 10): A meta-analysis. *Journal of Cognitive Neuroscience*, *18*(6), 932–948.

Hadwin, J., & Perner, J. (1991). Pleased and surprised: Children's cognitive theory of emotion. *British Journal of Developmental Psychology*, *9*, 215–234.

Koechlin, E., Ody, C., & Kouneiher, F. (2003). The architecture of cognitive control in the human prefrontal cortex. *Science*, *302*(5648), 1181–1185.

Koechlin, E., & Summerfield, C. (2007). An information theoretical approach to prefrontal executive

function. *Trends in Cognitive Science*, *11*(6), 229–235.

Leslie, A. M. (1987). Pretense and representation in infancy: The origins of 'theory of mind'. *Psychological Review*, *94*, 412–426.

Lohmann, G., Muller, K., Bosch, V., Mentzel, H., Hessler, S., Chen, L., et al. (2001). LIPSIA: A new software system for the evaluation of functional magnetic resonance images of the human brain. *Computerized Medical Imaging and Graphics*, *25*(6), 449–457.

Lohmann, G., Neumann, J., Mueller, K., Lepsien, J., & Turner, R. (2008, September). The multiple comparison problem in fMRI: A new method based on anatomical priors. MICCAI 2008, Workshop on Analysis of Functional Medical Images, New York.

Maguire, E. A. (2001). Neuroimaging studies of autobiographical event memory. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *356*(1413), 1441–1451.

Mitchell, J. P. (2008). Activity in right temporo-parietal junction is not selective for theory-of-mind. *Cerebral Cortex*, *18*(2), 262–271.

Mitchell, J. P., Banaji, M. R., & Macrae, C. N. (2005). General and specific contributions of the medial prefrontal cortex to knowledge about mental states. *NeuroImage*, *28*(4), 757–762.

Muller, N. G., & Knight, R. T. (2006). The functional neuroanatomy of working memory: Contributions of human brain lesion studies. *Neuroscience*, *139*(1), 51–58.

Norris, D. G. (2000). Reduced power multislice MDEFT imaging. *Journal of Magnetic Resonance Imaging*, *11*(4), 445–451.

Perner, J. (1993). *Understanding the representational mind*. Cambridge, MA: MIT Press.

Perner, J., Aichhorn, M., Kronbichler, M., Staffen, W., & Ladurner, G. (2006). Thinking of mental and other representations: The roles of left and right temporo-parietal junction. *Social Neuroscience*, *1*(3 & 4), 245–258.

Perry, J. (1994). Intentionality. In S. Guttenplan (Ed.), *A companion volume to the philosophy of mind* (pp. 386–395). Oxford, UK: Blackwell.

Petrides, M. (2005). Lateral prefrontal cortex: Architectonic and functional organization. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *360*(1456), 781–795.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral Brain Sciences*, *1*, 515–526.

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., et al. (2001). A default mode of brain function. *Proceedings of the National Academy of Science USA*, *98*, 676–682.

Rakcozy, H., Warneken, F., & Tomasello, M. (2007). "This way!", "No! That way!": 3-year olds know that two people can have mutually incompatible desires. *Cognitive Development*, *22*, 47–68.

Repacholi, B. M., & Gopnik, A. (1997). Early reasoning about desires: Evidence from 14- and 18-month-olds. *Developmental Psychology*, *33*(1), 12–21.

Ruff, C. C., Knauff, M., Fangmeier, T., & Spreer, J. (2003). Reasoning and working memory: Common

and distinct neuronal processes. *Neuropsychologia*, *41*(9), 1241–1253.

Samson, D., Apperly, I. A., Kathirgamanathan, U., & Humphreys, G. W. (2005). Seeing it my way: A case of a selective deficit in inhibiting self-perspective. *Brain*, *128*(5), 1102–1111.

Saxe, R. (2006). Uniquely human social cognition. *Current Opinion in Neurobiology*, *16*(2), 235–239.

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

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people: The role of the temporo-parietal junction in "theory of mind". *NeuroImage*, *19*(4), 1835–1842.

Saxe, R., & Powell, L. J. (2006). It's the thought that counts: Specific brain regions for one component of theory of mind. *Psychological Science*, *17*(8), 692–699.

Saxe, R., & Wexler, A. (2005). Making sense of another mind: The role of the right temporo-parietal junction. *Neuropsychologia*, *43*(10), 1391–1399.

Saxe, R., Xiao, D. K., Kovacs, G., Perrett, D. I., & Kanwisher, N. (2004). A region of right posterior superior temporal sulcus responds to observed intentional actions. *Neuropsychologia*, *42*(11), 1435–1446.

Searle, J. (1983). *Intentionality: An essay in the philosophy of mind*. New York: Cambridge University Press.

Shamay-Tsoory, S. G., & Aharon-Peretz, J. (2007). Dissociable prefrontal networks for cognitive and affective theory of mind: A lesion study. *Neuropsychologia*, *45*(13), 3054–3067.

Sommer, M., Dohnel, K., Sodian, B., Meinhardt, J., Thoermer, C., & Hajak, G. (2007). Neural correlates of true and false belief reasoning. *NeuroImage*, *35*(3), 1378–1384.

Spiers, H. J., & Maguire, E. A. (2006). Spontaneous mentalizing during an interactive real world task: An fMRI study. *Neuropsychologia*, *44*(10), 1674–1682.

Stringaris, A. K., Medford, N., Giora, R., Giampietro, V. C., Brammer, M. J., & David, A. S. (2006). How metaphors influence semantic relatedness judgments: The role of the right frontal cortex. *NeuroImage*, *33*(2), 784–793.

Svoboda, E., McKinnon, M. C., & Levine, B. (2006). The functional neuroanatomy of autobiographical memory: a meta-analysis. *Neuropsychologia*, *44*(12), 2189–2208.

Symons, C. S., & Johnson, B. T. (1997). The self-reference effect in memory: A meta-analysis. *Psychological Bulletin*, *121*, 371–394.

Talairach, J., & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain*. New York: Thieme.

Ugurbil, K., Garwood, M., Ellermann, J., Hendrich, K., Hinke, R., Hu, X., et al. (1993). Imaging at high magnetic fields: Initial experiences at 4 T. *Magnetic Resonance Quarterly*, *9*(4), 259–277.

Vogeley, K., Bussfeld, P., Newen, A., Herrmann, S., Happe, F., Falkai, P., et al. (2001). Mind reading:

Neural mechanisms of theory of mind and self-perspective. *NeuroImage*, *14*(1), 170–181.

Vogt, B. A., & Laureys, S. (2005). Posterior cingulate, precuneal and retrosplenial cortices: Cytology and components of the neural network correlates of consciousness. *Progress in Brain Research*, *150*, 205–217.

Wagner, A. D., Shannon, B. J., Kahn, I., & Buckner, R. L. (2005). Parietal lobe contributions to episodic memory retrieval. *Trends in Cognitive Science*, *9*(9), 445–453.

Weissman, D. H., Perkins, A. S., & Woldorff, M. G. (2007). Cognitive control in social situations: A role for the dorsolateral prefrontal cortex. *NeuroImage*, *40*(2), 955–962.

Wellman, H. M., & Bartsch, K. (1988). Young children's reasoning about beliefs. *Cognition*, *30*, 239–277.

Wheatley, T., Milleville, S. C., & Martin, A. (2007). Understanding animate agents: Distinct roles for the social network and mirror system. *Psychological Science*, *18*, 469–474.

Worsley, K. J., & Friston, K. J. (1995). Analysis of fMRI time-series revisited—again. *NeuroImage*, *2*(3), 173–181.

Zahn, R., Moll, J., Krueger, F., Huey, E. D., Garrido, G., & Grafman, J. (2007). Social concepts are represented in the superior anterior temporal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(15), 6430–6435.