Functional Activity and Effective Connectivity of the Posterior Medial Prefrontal Cortex During Processing of Incongruent Mental States

Tobias Schuwerk,[1,2]* Katrin Döhnel,[2] Beate Sodian,[1] Ingo R. Keck,[3] Rainer Rupprecht,[2] and Monika Sommer[2]

[1]Department of Psychology, Ludwig-Maximilians-University, Munich, Germany
[2]Department of Psychiatry and Psychotherapy, University of Regensburg, Germany
[3]Department of Psychology, University of Regensburg, Germany

◆ ━━━━━━━━━━━━━━━━━━━━━━━━━━━ ◆

Abstract: The neurocognitive components of Theory of Mind reasoning remain poorly understood. In particular the role of the posterior medial prefrontal cortex in the processing of other's mental states such as beliefs that are incongruent with one's own knowledge of reality is not clear-cut. It is unknown whether this region is involved in computing discrepant mental states or in subsequently resolving a response conflict between the discrepant others' and one's own beliefs. To test this, we adapted a false belief paradigm for the separate inspection of functional brain activity related to (1) the computation of diverging beliefs and (2) the subsequent consideration and selection of another's or one's own belief. Based on statistical parametric findings from functional neuroimaging, we employed dynamic causal modelling combined with Bayesian model selection to further characterize the interplay of resulting brain regions. In the initial computation of diverging beliefs, the posterior medial prefrontal cortex (pMPFC) and the bilateral temporoparietal cortex were crucially involved. The findings suggest that the bilateral temporal cortex engages in the construction and adjustment of diverging mental states by encoding relevant environmental information. The pMPFC inhibits this stimulus-bound processing which helps to compute discrepant mental states and process another's false belief decoupled from one's own perception of reality. In the subsequent question phase the right temporoparietal cortex showed increased activity related to switching to and reconsidering another's beliefs in order to select the correct response. *Hum Brain Mapp 35:2950–2965, 2014.* © 2013 Wiley Periodicals, Inc.

Key words: theory of mind reasoning; posterior medial prefrontal cortex; temporoparietal cortex; decoupling; false belief task; effective connectivity; dynamic causal modelling

◆ ━━━━━━━━━━━━━━━━━━━━━━━━━━━ ◆

INTRODUCTION

Reasoning about others' and one's own beliefs is a sophisticated cognitive ability. This form of Theory of Mind (ToM) reasoning becomes especially relevant when another's and one's own beliefs about a situation differ. For example, in Shakespeare's Romeo and Juliet, we only understand Romeo's fatal reaction at the end of the play if we understand that his behavior was based on the false belief that Juliet had died. Although real life situations are usually less dramatic, processing mental states that differ from our own is essential for our everyday social life.

For the investigation of the neural underpinnings of processing incongruent mental states, two sources of evidence are important. On the one hand, studies that aimed to specify the neural processing of beliefs by using classical false belief tasks. On the other hand, neuroscientific studies on visual perspective taking that addressed processing of another's viewpoint which does not correspond to one's own viewpoint in a certain situation.

Neuroimaging studies that investigated belief reasoning by using verbal stories or cartoons describing behavior based on another's belief revealed the involvement of the bilateral temporoparietal cortex, including the temporoparietal junction (TPJ), the posterior superior temporal sulcus (pSTS) and adjacent middle temporal gyrus [MTG; Saxe, 2009; Döhnel et al., 2012; van der Meer et al., 2011]. It is suggested that the temporoparietal cortex may be necessary to maintain and switch between one's own and another person's representation [cf., Hartwright et al., 2012], thereby requiring attentional orienting from stimulus-bound processing (e.g., reasoning about the state of reality) to stimulus-independent processing [reasoning about unobservable beliefs of others; Döhnel et al., 2012].

Currently, the role of the medial prefrontal cortex (MPFC) in ToM reasoning remains unclear. Interpretations range from the claim that the MPFC is a key region in ToM reasoning [e.g., Frith and Frith, 2003] to the view that it plays no crucial role [cf., Bird et al., 2004; Aichhorn et al., 2009]. Based on findings from the first neuroimaging studies on ToM reasoning, Gallagher and Frith [2003] suggested that the MPFC underpins the decoupling mechanism necessary to reason about another's mental state that does not correspond to one's own perception of reality. Contrary to this, unimpaired ToM reasoning in a patient with a medial frontal lobe lesion led to the conclusion that this region may not be necessary for ToM reasoning at all [Bird et al., 2004]. A third viewpoint postulates that activity of the MPFC may be associated with ToM tasks, but that this activity is not specific for ToM reasoning and rather related to more general processing of information about others, including for example their physical appearance or internal states such as being hungry [Saxe and Powell, 2006; see also Saxe, 2009].

Studies that specifically investigated the neural correlates of false belief reasoning by using the object transfer task [originally from Wimmer and Perner, 1983] suggest that the posterior part of the MPFC (pMPFC, also referred to as dorsal MPFC; extending into the dorsal anterior cingulate cortex, dACC) might be involved in the decoupling mechanism [Rothmayr et al., 2011; Sommer et al., 2007; cf., van der Meer et al., 2011; Döhnel et al., 2012]. The object transfer task requires processing another's false belief about an object's location. A critical feature here is that the false belief of the protagonist about the object's location does not match with the participant's knowledge about the actual location of the object. This means the participant has to process someone's belief independently from his or her own knowledge. Alternative false belief tasks using for example verbal vignettes, contain a character's false belief, but do not involve the participant's own perception of this situation [cf., Saxe and Kanwisher, 2003; Bird et al., 2004; Aichhorn et al., 2009]. Thus, in contrast to that, the object transfer task saliently elicits a conflict between another's and one's own belief about the same situation. The pMPFC might engage in the processing of those conflicting mental states.

However, the pMPFC's role in false belief reasoning in the object transfer task is not clear-cut. These tasks not only require the computation of another's belief that is incongruent to one's own knowledge about a situation. Another aspect is to consider and select the correct belief in order to give a response. The other's and one's own mental state compete for response selection. Therefore it is possible that the pMPFC activity observed in object transfer tasks may be associated with cognitive control processes in response selection [cf., Botvinick et al., 2004; Ridderinkhof et al., 2004; Venkatraman et al., 2009] and not with belief computation or the decoupling mechanism.

In the domain of visual-perspective taking, such a distinction between a computation and a selection process was shown [cf., Leslie et al., 2005; Qureshi et al., 2010]. Derived from findings of a visual-perspective taking paradigm, two distinct phases and processes were suggested: (1) in a computation phase another's and one's own viewpoint are computed and represented. This process may be linked to activity in the bilateral temporoparietal cortex [McCleery et al., 2011]. (2) In an adjacent selection phase, inhibitory processes related to activity in lateral frontal cortices are proposed to prioritize and select one viewpoint over the other [McCleery et al., 2011; Ramsey et al., 2013].

These findings raise the question of whether functional activity related to false belief reasoning can be specified with respect to a belief computation and a belief selection phase. Such a distinction would be of particular interest for pMPFC activity: is it associated with the computation of another's false belief, presumably requiring a decoupling mechanism, or is it involved in dealing with a response conflict and response selection, resulting from the discrepancy between one's own and anothers' belief. As assumptions about such a distinction can so far only be based on findings from visual-perspective taking paradigms, conclusions to the processing of diverging mental states such as beliefs remain tentative.

In the present functional resonance imaging (fMRI) study, we adapted a classical object transfer false belief paradigm [Baron-Cohen et al., 1985] in a way that reasoning about one's own and another's belief could be conceptually separated into two phases: (1) in an initial story phase, another's belief about an object's location had to be computed. It was either congruent or incongruent to one's own knowledge about the situation. (2) In the following question phase, one's own or the other's belief, either corresponding or diverging, had to be considered and selected for response. In this manner we explicitly

separated the processing of emerging discrepant beliefs from the need to select one belief over the other.

We predicted that if the pMPFC is crucial for computing incongruent beliefs, activity in this region should be observed in the story phase when diverging beliefs had to be processed. If the pMPFC is necessary to resolve a response conflict to select among competing mental states, this region should be more active in the question phase when incongruent beliefs are already computed and one has to be selected for response.

We were further interested in functional activity of the bilateral temporoparietal cortex with respect to the computation and selection phase. Prior literature holds arguments for both phases: McCleery et al. [2011] argued that this region is involved in the computation and representation of visual perspectives. However, studies on neural underpinnings of belief reasoning concluded that this region might be involved in maintaining and switching between one's own and another's representations [cf., Döhnel et al., 2012; Hartwright et al., 2012]. This process is also relevant in the selection of the respective mental state. We hypothesized that if the bilateral temporoparietal cortex is related to the computation of beliefs, this region would be active in the story phase. If the bilateral temporoparietal cortex is involved in maintaining and switching between mental states in order to select one mental state over another, this region should show increased functional activity in the question phase of the current paradigm.

A second aim of this study was to shed light on the interaction of the pMPFC and temporoparietal areas involved in false belief reasoning. Although a huge body of research reports a remarkably consistent set of brain regions [see Mar, 2011], the interplay of those regions in ToM reasoning remains poorly understood. Based on findings from conventional statistical parametric mapping (SPM) of the current paradigm, we employed dynamic causal modelling (DCM) combined with Bayesian model selection (BMS) to characterize the effective connectivity of resulting brain regions. Of particular interest was how those brain regions interact when incongruent beliefs are processed.

METHODS

Participants

Twenty-one subjects (mean age 23.3 years, SD = 3.7 years; 12 females) took part in the study. All participants were right-handed, had no neurological or psychiatric history and gave informed written consent. The University Medical Center Regensburg ethics committee approved the study.

Task and Procedure

The present task was based on the Sally-Anne paradigm by Baron-Cohen et al. [1985]. To isolate the difference between the protagonist's and one's own mental state, we varied the following aspects compared to other false belief tasks: first, instead of two characters, only one character was used in the story. Second, in addition to asking for the other's false belief, we asked the subjects for their own belief. In this way the participants were encouraged to process the other's mental state in contrast to their own mental state. Each trial consisted of two phases (Fig. 1A): in an initial story phase the character formed a belief that was either incongruent or congruent with the participant's knowledge about the situation (incongruent-beliefs condition vs. congruent-beliefs condition). In a subsequent question phase, subjects were either asked for the character's belief (other-incongruent vs. other-congruent) or their own belief (self-incongruent condition vs. self-congruent condition). Cartoon still images and animated videos were prepared using Adobe CS5.5 (Adobe Systems, San Jose, CA).

The story phase depicted a room with a boy standing on a wooden floor. A dark-brown and a light-brown box were on his left and right hand side, respectively. A basement was underneath the wooden floor. Subjects were instructed that what happens in this basement during the video clip can only be seen by them, but not by the character. In the beginning of the video a ball fell from the ceiling into one of the two open boxes and then the boxes closed. After the boxes closed, two events simultaneously took place. One event, which was not observed by the character, involved the ball bouncing into the basement through a hidden trap door. The other event, which was observed by the character, was that the boxes switched places. In the congruent-beliefs condition, the ball bounced into the basement, bounced right back into the same box where it had been before, and was transferred inside this box to the character's other side. Thus, at the end of the video both the character and the subject had the same belief about the ball's location. In the incongruent-beliefs condition, the ball also bounced into the basement, but bounced back with a short time delay so that it did not take part in the switch of the two boxes and thus re-entered through the trap-door into the other box. In this condition the character ends up with a false belief about the objects' location (assuming the ball was transferred with the box it initially entered), diverging from the subject's belief (knowing the ball was not transferred and is now in the other box).

The video lasted for 4 s and was followed by a time-varying fixation picture (2–4 s, Fig. 1B) which contained the same scene including the two closed boxes. The character was replaced by a fixation cross. In the subsequent question picture (2 s, Fig. 1C), the fixation cross disappeared and the participants had to reason in which one of the two boxes the story character thinks the ball is (indicated by the German word "Er?", English: "He?") or in which box the participant thinks the ball is (indicated by the German word "Ich?", English: "I?"). Depending on the video presented previously, reasoning about the character's false belief (other-incongruent condition) or true
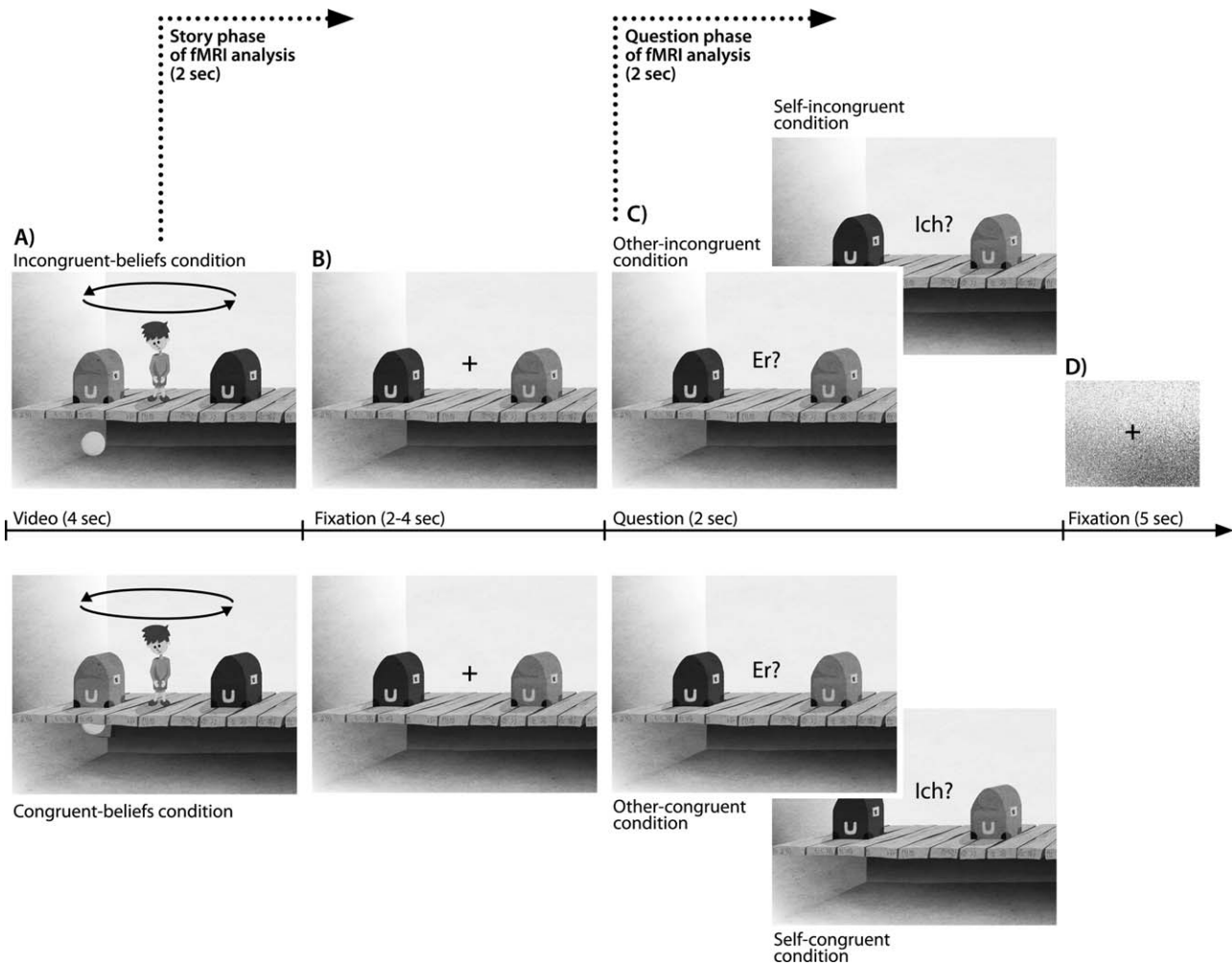
Figure 1.

Example trial and experimental design. (**A**) Story phase: A trial started with the presentation of a 4-s long video, in which the character either ended up with an incongruent belief (incongruent-beliefs condition), or with a congruent belief (congruent-beliefs condition) compared to one's own belief about the ball's location (**B**) Fixation picture with jittered duration (2–4 s) (**C**) Question phase: In this picture the subject had to respond either according to the other's belief ("Er?"), or the own belief ("Ich?") about the location of the ball. Depending on the prior presented video the beliefs could either diverge or be congruent. Response was given via button press (left or right) (**D**) The trial ended with a second fixation picture, lasting for 5 s. Trials from each condition were presented in an interleaved fashion.

belief (other-congruent condition) was required. When the subjects were asked for their own belief, it either diverged from the character's belief (self-incongruent condition) or corresponded with it (self-congruent condition). Participants were instructed to respond as fast and as accurately as possible. Reaction time was measured from the onset of the question picture until a left or right response button was pressed (LUMItouch optical response device; Photon Control Inc., Burnaby, Canada). The trial ended with a second fixation picture (5 s, Fig. 1D) showing a fixation cross underplayed with scrambled pixels, matched to the prior stimuli in color, contrast and luminance.

To prevent habituation, the experimental trials were intermixed with filler trials. The subjects were either asked which of the boxes was light-brown (in German: "hell?"; English: "light-brown?") or which of the boxes was dark-brown (in German: "dunkel?"; English: "dark-brown?"). These filler trials were included to increase task demands and avoid the predictability of the questions.

A total of 180 trials (30 trials per condition and 60 filler trials) were presented in pseudo-randomized order. The videos were controlled for type, number, order and laterality of events. Question pictures were identical across conditions, showing solely the two closed boxes in the scene

without social stimuli (boy was absent). Prior to testing, subjects received a standardized instruction and completed several test trials outside the scanner until they were familiar with the stimuli. The software Presentation (Neurobehavioral Systems, Albany, CA) was used to present the stimuli and record responses. The stimuli were projected onto a screen in the MRI that could be seen through a mirror attached to the head coil. The subjects' head was kept in position via foam padding.

Behavioral Data Analyses

We performed a 2×2 repeated measures analysis of variance (ANOVA) using IBM SPSS statistics 20 (SPSS Inc., Chicago, IL). Within subject variables, considered person (other vs. self) and congruency of beliefs (incongruent vs. congruent) were considered. Reaction times (RTs) of correctly answered trials and response accuracy (percentage of correct responses) were analyzed. The significance level for the analyses was set at $P \leq 0.05$.

Imaging and Image Analyses

Image acquisition

Scanning was performed on a 3-Tesla head scanner (Siemens Allegra, Erlangen, Germany). Blood oxygenation level dependent (BOLD) functional MRI was measured using a T_2^*-weighted echo-planar imaging (EPI) sequence (TR = 2 s, TE = 0.05 s, 90° flip angle, FoV = 192 mm, plane matrix 64×64, voxel size $3 \times 3 \times 3$ mm^3). Each EPI volume contained 32 axial slices. A total of 1,275 volumes was acquired. Structural images included a T_1^*-weighted MPRAGE (Magnetization Prepared Rapid Acquisition Gradient Echo) sequence (TR = 2.25 s, TE = 0.026 s, TI = 0.9 s, FoV = 256 mm, voxel size $1 \times 1 \times 1$ mm^3, 160 axial slices). The entire scan session lasted approximately 45 min.

Preprocessing and statistical analyses

All images were preprocessed and analyzed with SPM8 (Wellcome Department of Imaging Neuroscience, London, UK), run in MATLAB 7 (The MathWorks, Natick, MA). Differences in acquisition time between functional slices were compensated by slice-time correction using the middle slice as a reference. Images were spatially realigned to the first volume by rigid body transformation to correct for head movements. After coregistration to the structural T_1^*-weighted images, data were normalized to the Montreal Neurological Institute (MNI) standard space with a voxel size of $2 \times 2 \times 2$ mm^3. Images were spatially smoothed with a 8 mm full-width-at-half-maximum Gaussian kernel.

In this event-related design, all statistical analyses were based on functional activity obtained from the whole brain. A first-level fixed effects analysis was computed for each subject with the general linear model [GLM; Friston, 1995]. The analysis focused on amplitude changes in the hemody-

namic response function (HRF) associated with the cognitive processes of interest. For each condition correctly answered, trials were modeled as a boxcar function with a 2 s interval convolved with the HRF [Friston, 1998]. In the story phase the two regressors for the incongruent-beliefs condition and the congruent-beliefs condition comprised the last 2 s of the 4-s long video. The events in the video were timed so that exactly after 2 s it became clear whether the characters' belief was incongruent or congruent to one's own belief. In the question phase, which started with the onset of the question and lasted for 2 s, we modeled a regressor for each of the four conditions of interest (other-incongruent condition, other-congruent condition, self-incongruent condition and self-congruent condition) and the filler trials. Realignment parameters, a single covariate representing the mean (constant) over scans, and a non-hits parameter (incorrect responses and misses) were included as regressors of no interest. The data were high-pass filtered with a frequency cutoff at 128 s.

Statistical parametric maps (SPMs) were generated for each subject by *t*-statistics derived from contrasts utilizing the HRF [Friston et al., 2002]. To identify brain regions associated with the processing of incongruent versus congruent beliefs in the initial story phase, we contrasted the conditions incongruent-beliefs versus congruent-beliefs.

In order to detect brain activity related to considering and selecting incongruent beliefs that elicit a response conflict in the subsequent question phase, we analyzed the main effect of congruency of beliefs [(other-incongruent + self-incongruent) versus (other-congruent + self-congruent)]. Additionally, the main effect considered person was calculated [(other-incongruent + other-congruent) versus (self-incongruent + self-congruent)]. We were further interested in an interaction effect between the considered person and congruency of beliefs [(other-incongruent - other-congruent) > (self-incongruent–self-congruent)] and vice versa.

For group analyses, the single-subjects' first-level contrasts were taken to a second-level random effects analysis. To explore whether common brain regions are associated with processing incongruent beliefs in the story phase and considering another's beliefs in the question phase, we computed a conjunction analysis [based on the Minimum statistic compared to the Null Conjunction; Nichols et al., 2005] on the contrasts (incongruent beliefs > congruent beliefs) and [(other-incongruent + other-congruent) > (self-incongruent + self-congruent)]. The resulting set of significant voxel values for each contrast constituted a SPM map. The SPM maps were thresholded at $P \leq 0.001$, uncorrected. Reported significant voxels survived a statistical threshold of $P < 0.001$, FWE (family-wise error)-corrected for multiple comparisons on cluster level. Approximate anatomical labels were assigned to significantly active regions using the Anatomical Automatic Labeling toolbox [as described by Tzourio-Mazoyer et al., 2002] for SPM. MRIcron (www.mccauslandcenter.sc.edu/mricro/mricron/) was used to visualize resulting activation maps.

Dynamic Causal Modelling

To characterize effective connectivity from the observed fMRI data, DCM10 implemented in SPM8 was used. DCM allows inferring unobserved neuronal states from the observed BOLD response [Friston et al., 2003; Stephan et al., 2007]. Bilinear dynamic causal models (DCMs) can clarify how a brain region influences another and how this is modulated by task demands [cf., Friston et al., 2003; Stephan et al., 2010]. The following parameters can be specified: (1) fixed connections among neuronal populations reflect the effective connectivity in the experiment; (2) driving inputs describe the direct influence of experimental stimulation on specific regions and (3) modulatory inputs describe the change of strength of connectivity between those neuronal populations, depending on experimental conditions. Resulting neurodynamic and hemodynamic parameters are estimated by predicting BOLD signals with optimal fit to the observed BOLD signals. Subsequently the estimated models can be compared using Bayesian model selection (BMS) to select the model that is most likely among the other defined models, given the data. This method balances model fit and model complexity [Stephan et al., 2009]. For the winning model. inferences about its parameters can be made [cf., Stephan et al., 2010].

Selection of brain regions and time series extraction

We investigated the connectivity profile of three brain regions: left and right temporoparietal cortex (left and right pSTS/MTG) and pMPFC. Volumes of interest (VOIs) were defined for each subject based on the following considerations: (1) these regions showed the most robustly significant increased activity on the group level in the contrast incongruent-beliefs > congruent-beliefs from the story phase. (2) A number of previous meta-analyses suggested that these regions play a crucial role in ToM reasoning [Gobbini et al., 2007; van Overwalle and Baetens, 2009; van Overwalle, 2009; Mar, 2011]. Furthermore a recent analysis of effective connectivity during perspective taking reported increased connectivity between those temporoparietal regions and the MPFC in a social context [Hillebrandt et al., 2013].

As the locations of activated brain regions vary across subjects, functional and anatomical inclusion criteria were defined to make models comparable across subjects [Stephan et al., 2010; cf. Vossel et al., 2012; Ma et al., 2012]: Time series from local maxima significantly activated in single-subject analysis ($P < 0.05$, uncorrected on cluster level) were extracted which were closest to the corresponding group maximum (for one right pSTS/MTG and two pMPFC clusters that did not reach this level of significance, the corresponding group coordinates were used). We further ensured that single-subject activations be within the same anatomical area as the group maximum, as defined by the Anatomical Automatic Labeling atlas [Tzourio-Mazoyer et al., 2002]. The mean MNI coordinates (x,y,z; $\pm$SD) for the three regions were: left pSTS/MTG,

-53 ± 5.7, -58 ± 14, 12 ± 5.7; right pSTS/MTG, 51 ± 8.4, -68 ± 13, 9 ± 7; pMPFC, -3 ± 6.5, 12 ± 21, 62 ± 15. The first eigenvariate of activated voxels within 15 mm of subject-specific maxima was extracted. In doing this we addressed the difficulty of integrating large cluster sizes of peak activations on the group-level and having high inter-individual variability of the locations of significant activation while still accounting for comparability across subjects [cf., Stephan et al., 2010; DiQuattro and Geng, 2011]. In three subjects activations did not meet these criteria. These subjects were excluded so that the final DCM analysis was performed on the remaining 18 participants.

Definition of fixed connections

Sparse evidence of the connectivity of human brain regions challenges the definition of fixed connections especially in DCMs of higher cognitive processes [Penny et al., 2004]. Evidence from diffusion-weighted imaging tractography and resting-state functional connectivity suggests that regions of the posterior temporal/parietal cortex are connected to the medial prefrontal cortex [cf., Mars et al., 2011; Hagmann et al., 2008]. It was shown that structural connections between posterior temporal cortices and medial prefrontal areas in the macaque brain are reciprocal [Barbas et al., 1999]. Based on these findings we specified bidirectional endogenous connections between left pSTS/MTG and pMPFC (left pSTS/MTG → pMPFC, pMPFC → left pSTS/MTG), as well as between right pSTS/MTG and pMPFC (right pSTS/MTG → pMPFC, pMPFC → right pSTS/MTG). Structural brain analyses [Hagmann et al., 2008] support the notion that brain networks within temporal and posterior parietal cortical regions are indirectly linked via connector hubs such as the precuneus. We thus did not assume direct connections between left and right pSTS/MTG in our DCMs.

Definition of driving inputs

Human as well as primate studies suggest that middle temporal gyrus and posterior temporal regions integrate information from multisensory input [Beauchamp et al., 2004; Barnes and Pandya, 1992] and play a role in semantic processing of this information [Visser et al., 2012] or action observation [cf., Frith and Frith, 2003; Caspers et al., 2010], for example. Given these findings we modeled experimental stimulation from the story phase as driving inputs to left and right pSTS/MTG. Because no a priori assumption could be derived whether left pSTS/MTG, right pSTS/MTG or both regions would receive exogenous input, we defined three input families for each possibility (left pSTS/MTG, right pSTS/MTG, BOTH).

Definition of modulatory inputs

Of particular interest was if and how effective connectivity changes during the incongruent-beliefs condition. In

TABLE I. Mean reaction time and response accuracy in the question phase, listed for each condition

Condition	RT (ms)		PCR (%)	
	M	SD	M	SD
other-incongruent	838	240	97	4.3
other-congruent	726	197	99	1.7
self-incongruent	861	226	95	5.3
self-congruent	718	187	98	2.3

Notes: RT, reaction time; PCR, percentage of correct responses; M, mean; SD, standard deviation.

other words, how do those regions interact when incongruent beliefs have to be processed? Thus within each input family, models with all possible variations of modulatory input from the incongruent-beliefs condition were specified. We allowed incongruent trials to influence connection strength of either one, two, three or all four of the above defined fixed connections. All possible combinations were varied. This approach resulted in 48 different DCMs (16 DCMs per input family) which were fitted for each subject.

Bayesian model selection

To identify the optimal model we used random-effects BMS to account for variability of effects across subjects [Penny et al., 2004; Stephan et al., 2009, 2010]. This method was applied in two steps: in order to reduce uncertainty about driving input regions, we initially compared the three defined input families [Penny et al., 2010; Stephan et al., 2010]. Subsequently the models within the input family with the highest posterior exceedance probability were compared to determine the most likely DCM. To construct a model that represents the winning model from group-level BMS, we employed Bayesian parameter averaging (BPA). By computing a joint posterior density across all subjects, this method provides averaged parameters which are representative of parameters from the individual winning model [cf., Garrido et al., 2007; Acs and Greenlee, 2008]. This method weighs the influence of individual subjects according to their within-subject variance. In this way it is more robust against outliers compared to classical approaches based on, for instance, t statistics [cf., Neumann and Lohmann, 2003].

RESULTS

Behavioral Results

Mean RT and response accuracy for each experimental condition are shown in Table I. Regarding RT, the ANOVA revealed a significant main effect of congruency of beliefs, $F(1,20) = 30.27$, $P < 0.001$, $\eta_p^2 = 0.602$. RTs were slower when incongruent beliefs ($M = 849$ ms) had to be considered, compared to considering congruent beliefs

($M = 722$ ms), regardless of whether one's own or another's belief was requested. No effect of considered person, $F(1,20) = 0.34$, $P = 0.56$, $\eta_p^2 = 0.017$, and no significant interaction between considered person and congruency of beliefs, $F(1,20) = 1.94$, $P = 0.18$, $\eta_p^2 = 0.088$, was observed. The same pattern of results was observed for response accuracy. The ANOVA showed a significant main effect for congruency of beliefs, $F(1,20) = 8.41$, $P = 0.009$, $\eta_p^2 = 0.296$. Participants gave more correct responses when they were asked for congruent ($M = 98.3\%$) compared to incongruent beliefs ($M = 95.9\%$). Neither the main effect of the considered person, $F(1,20) = 1.23$, $P = 0.28$, $\eta_p^2 = 0.058$, nor the interaction between the considered person and congruency of beliefs, $F(1,20) = 0.28$, $P = 0.60$, $\eta_p^2 = 0.014$, were significant.

Whole Brain Imaging Results

Story phase

In this phase of the trial, it became clear for the participants whether the character's belief diverged from or corresponded to their own belief. However, the subjects did not yet know whether their own belief or that of the other's would be asked about in the subsequent question phase. In order to determine which brain regions are engaged in the processing of emerging incongruent beliefs, we contrasted the incongruent-beliefs with the congruent-beliefs condition (see Table II, Fig. 2A). The contrast revealed increased functional activity in the median/anterior cingulate region (pMPFC/dACC; BA32), left and right pSTS/MTG (BA19/37/39), left inferior frontal gyrus (IFG; BA6/44), right inferior area, right superior parietal gyrus (BA7), and left caudate nucleus and precental gyrus/frontal eye field (FEF; BA6). In the reverse contrast, congruent-beliefs condition over incongruent-beliefs condition, no brain region showed significantly increased activity.

Question phase

To identify brain regions related to processing a response conflict due to diverging mental states, we compared the conditions asking for an incongruent belief to conditions in which the character's belief and the subject's own belief about the situation were congruent. Results based on the full factorial design revealed no main effect of congruency of beliefs. Furthermore, no main effect of the considered person, or an interaction of a considered person × congruency of beliefs was observable. We next examined whether the direct t-contrast of conditions in which the subjects had to consider the other's belief compared to conditions in which the subjects had to respond according to their own belief would reveal significant functional activity. This allowed us to identify brain regions activated when considering another's compared to one's own mental states (Table II, Fig. 2B). On a lowered cutoff for FDR (false discovery rate)-corrected results

TABLE II. Significant clusters ($p_{\text{FWE-corr}} < .001$, cluster level) of functional activity from the two test phases: (1) story phase: peak activations associated with the processing of incongruent perspectives; (2) question phase: peak activation associated with the selection of the other's beliefs (other > self); (3) conjunction of story phase and question phase: peak activations commonly associated with the processing of incongruent perspectives and the selection of the other's beliefs

Contrast/brain region	Left/right	BA	MNI coordinates			Cluster size[a]	T-value[b]
			x	y	z		
(1) Story phase: incongruent beliefs > congruent beliefs							
Left temporoparietal cortex: Left pSTS/left MTG/ middle occipital gyrus	L	19/39	−46	−74	4	17325	16.85
Right temporoparietal cortex: Right pSTS/right MTG/middle occipital gyrus	R	37/39	50	−64	8		10.16
Superior parietal gyrus	R	7	18	−80	50		10.09
Right caudate nucleus	R	−	18	26	2	695	5.94
Anterior cingulum	R	−	22	36	0		4.19
Left caudate nucleus	L	−	−6	24	4		5.29
Precentral gyrus/FEF	L	6	−36	−8	54	773	5.71
			−46	−10	52		5.50
			−54	0	46		5.15
Median/anterior cingulate region/ pMPFC/dACC	L	32	−4	24	34	1592	5.69
	L	6/8	−4	10	58		5.68
	R	8/32	14	24	38		5.59
IFG[c]	L	6/44	−62	8	10	232	5.30
			−60	14	4		4.52
			−44	8	2		3.85
Inferior frontal area	R	−	24	4	30	646	5.10
			28	−6	58		5.02
			36	−10	62		4.98
(2) Question phase: other > self [(other-incongruent + other-congruent) > (self-incongruent + self-congruent)]							
Right temporoparietal cortex: Right pSTS/right MTG/ middle occipital gyrus[d]	R	39	44	−52	4	68	6.17
(3) Conjunction of story phase an question phase: (incongruent beliefs > congruent beliefs) and [(other-incongruent + other-congruent) > (self-incongruent + self-congruent)]							
Right temporoparietal cortex: Right pSTS/right MTG/ middle occipital gyrus[e]	R	39	44	−52	4	95	5.74

Notes: Labeling of Brodmann areas (BAs) is approximate. Brain regions highlighted in bold are of particular interest. pSTS, posterior superior temporal sulcus; MTG, middle temporal gyrus; FEF, frontal eye field; IFG, inferior frontal gyrus.
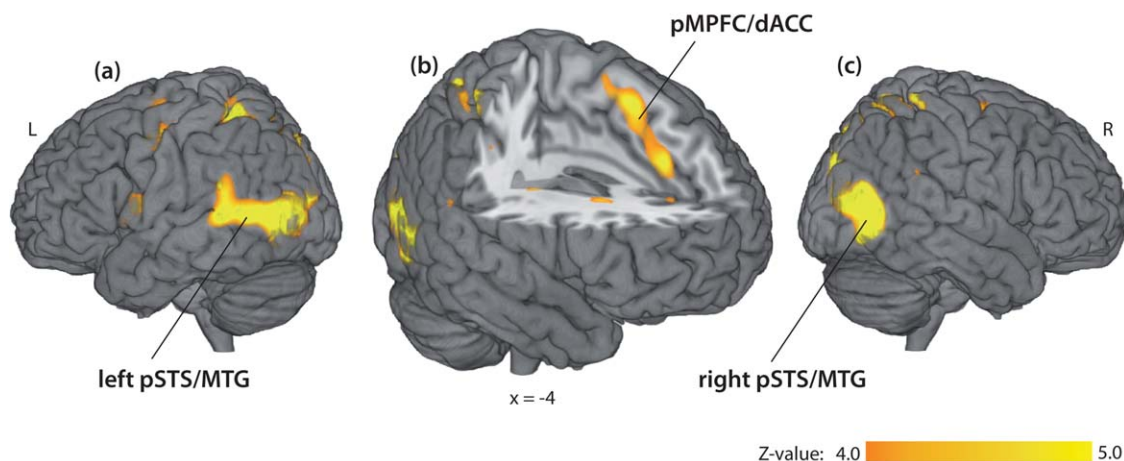[a]Number of activated voxels per cluster.
[b]Peak T-value in activated cluster.
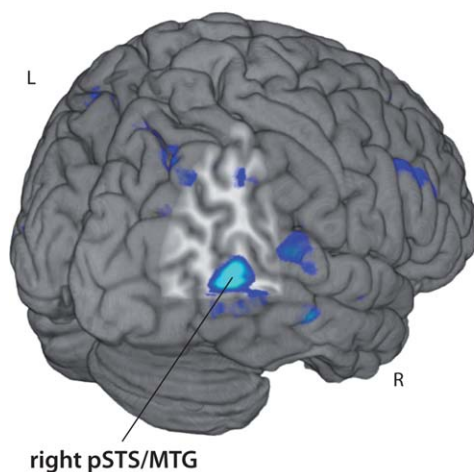[c]$p_{\text{FDR-corr}} = 0.004$, cluster-level.
[d]$p_{\text{FDR-corr}} = 0.23$, cluster-level.
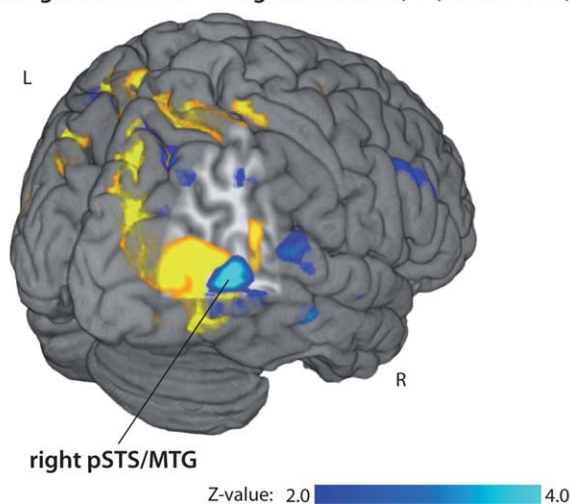[e]$p_{\text{FDR-corr}} = 0.096$, cluster-level.

Figure 2.

A: Processing of incongruent beliefs in the story phase: Whole-brain fMRI group findings. Activation maps for the contrast incongruent beliefs > congruent beliefs from the initial story phase. Colored regions indicate significantly activated voxels with $Z > 4.0$, $p_{FWE-corr} < 0.001$, cluster level. All results are overlayed onto a MRI brain template and are displayed in neurological convention (L and R indicate left and right hemispheres). (a) Rendered brain showing significantly activated voxels in the left posterior superior temporal sulcus (pSTS)/middle temporal gyrus (MTG). (b) Significantly increased functional activity of the posterior medial prefrontal cortex including dorsal anterior cingulate cortex (pMPFC/dACC). (c) Activity in the right pSTS/MTG rendered on the lateral surface of the brain. **B**: Considering other's beliefs in the question phase: Whole-brain fMRI group findings. Activation maps for the contrast other > self [(other-incongruent + other-congruent) versus (self-incongruent + self-congruent)]. Colored regions indicate significantly activated voxels with $Z > 2.0$, $p_{FDR-corr} = 0.23$, cluster level. Right temporoparietal cortex (right pSTS/MTG) showed increased functional activity when subjects had to respond according to the other's mental state, irrespective whether it was incongruent or congruent to the own belief. **C**: Activation maps from the story and the question phase, overlayed onto a single rendered brain: a part of the right temporoparietal cortex showed increased activity in the contrast incongruent beliefs > congruent beliefs from the story phase (indicated in yellow) and in the contrast other > self [(other-incongruent + other-congruent) versus (self-incongruent + self-congruent)] from the question phase (indicated in blue).
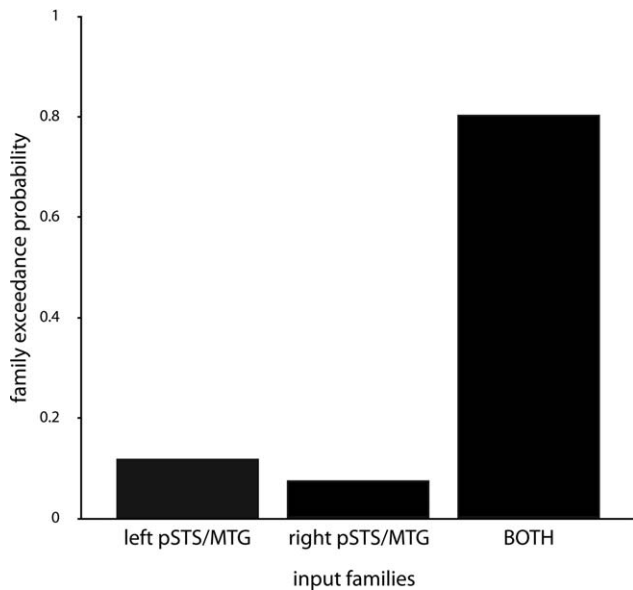
Figure 3.

Bayesian Model Selection (BMS) part 1: comparison of input families. Exceedance probabilities of each input family. The family of models that received input to bilateral temporoparietal cortex (left and right pSTS/MTG, BOTH) had the greatest evidence.

[$p_{FDR-corr} = 0.23$, cluster-level; cf., Burnett et al., 2009], increased functional activity in the right pSTS/MTG (BA 39) was observed for this contrast. In the reverse contrast no brain region showed increased functional activity.

To check whether there was brain activity in the question phase associated with the prior established context of incongruent beliefs we additionally contrasted the filler trials with incongruent context (incongruent beliefs in the prior story phase) with filler trials with congruent context (congruent beliefs in the prior story phase). In the question phase of the filler trials the subjects had to consider the colors of the boxes, regardless of the previous processing of the other's or their own belief. No brain region showed a significant increase in functional activity in this contrast. The context of incongruent compared to congruent beliefs was not associated with differential functional activity in nonmental reasoning in the filler trials. This indicates that reasoning about the physical reality in the response phase does not involve brain regions similar to those that were found in considering and selecting another person's beliefs.

Common activity in the story and the question phase

The whole brain conjunction analysis of the contrasts (incongruent beliefs > congruent beliefs) and [(other-incongruent + other-congruent) > (self-incongruent + self-congruent)] showed on a lowered cutoff for FDR-corrected results ($p_{FDR-corr} = 0.096$, cluster-level) commonly increased functional activity of a part of the right pSTS/MTG (BA 39) associated with both the proc-

essing of emerging incongruent beliefs in the story phase and considering the other's belief to give the correct response in the question phase.

DCM results

Whole-brain SPM analyses suggested an important role of the left and right pSTS/MTG and the pMPFC in processing diverging mental states: those regions showed the largest amount of activated voxels on the group level in the contrast incongruent-beliefs > congruent-beliefs from the story phase. DCM was used to specify the interplay between those regions in this contrast. Based on functional and structural criteria as well as evidence from previous meta-analyses on neural bases of ToM reasoning and literature on brain structure, we defined 48 plausible DCMs. This set was partitioned based on regions that received driving inputs (left pSTS/MTG, right pSTS/MTG, BOTH). A Bayesian model comparison was conducted in two steps. Resulting exceedance probabilities can be found in Supporting Information Table SI. In the first step, random-effects BMS was employed to choose the optimal input family. The family of models that received input to both regions (BOTH) had the highest evidence (exceedance probability $\Phi = 0.80$; Fig. 3). In the second step, the 16 models within this winning family were compared. Those models varied in the modulatory influence of incongruent beliefs on fixed connections between left pSTS/MTG, right pSTS/MTG and MPFC. In the random-effects BMS, one model clearly outperformed all other models (exceedance
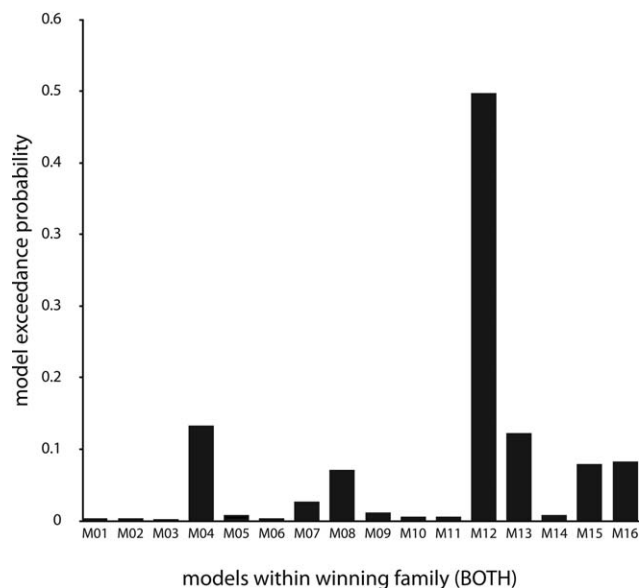

Figure 4.

Bayesian Model Selection (BMS) part 2: model comparison within winning family. Exceedance probabilities of the 16 models from the winning input family (BOTH). M12 clearly outperformed all other models.
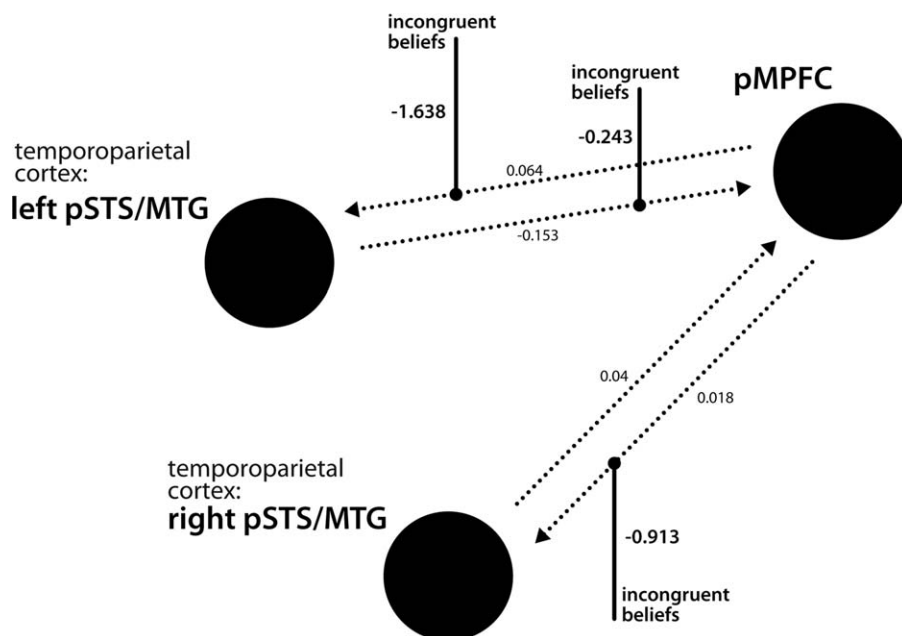
story phase: BPA results

Figure 5.

Story phase: Bayesian Parameter Averaging (BPA) results. Representative model averaged across winning models (M12) from each participant. Small numbers next to dotted lines indicate effective connectivity (endogenous parameters) between regions (left and right posterior superior temporal sulcus (pSTS)/middle temporal gyrus (MTG); posterior medial prefrontal cortex, pMPFC). Under the modulatory influence of the incongruent-beliefs condition (big bold numbers indicate modulatory parameters) the connections pMPFC→left pSTS/MTG and pMPFC→right pSTS/MTG became negative. An increased inhibitory effect was observed for the connection left pSTS/MTG→pMPFC.

probability $\Phi = 0.51$; model M12; Fig. 4). In this winning model, the incongruent beliefs affected connection strength of pMPFC→left pSTS/MTG, pMPFC→right pSTS/MTG and left pSTS/MTG→pMPFC.

A representative model averaged across winning models from each participant was built using BPA. The averaged parameters revealed a low effective connectivity between the modeled regions (see Supporting Information Table SII and Fig. 5). However, this changed under the modulatory influence of the incongruent-beliefs condition. Positive intrinsic connections pMPFC→left pSTS/MTG and pMPFC→right pSTS/MTG became negative under the influence of negative modulatory parameters. The negative intrinsic connection left pSTS/MTG→pMPFC became more negative through the modulatory effect (see Supporting Information Table SII and Fig. 5). Posterior probabilities yielded high confidence for all effects ($P(|\text{connection}| > 0.00) = 100\%$).

DISCUSSION

Reasoning about others' mental states, which do not correspond to one's own knowledge about the world is essen-

tial in social interaction and known as false belief reasoning [Frith and Frith, 2012]. This study aimed to investigate the role of the pMPFC for this important social cognitive ability. Of particular interest was whether the pMPFC is involved in computing incongruent beliefs or rather in resolving a response conflict due to another's belief that diverges from one's own. We thus adapted a false belief paradigm for the separate inspection of functional brain activity related to (1) the computation of diverging beliefs and (2) the subsequent consideration and selection of another's or one's own belief being either incongruent or congruent. In a second step we combined DCM with BMS to characterize the effective connectivity between the pMPFC and temporoparietal brain regions when incongruent beliefs are processed.

Computing Incongruent Beliefs (Story Phase)

By contrasting the incongruent beliefs condition with the congruent beliefs condition we intended to identify functional activity related to the initial computation of incongruent beliefs. When it became clear that another's belief was

diverging from one's own belief, activity was observable in the pMPFC/dACC which is in line with findings from earlier studies on false belief reasoning [e.g., Vogeley et al., 2001; Grèzes et al., 2004; Sommer et al., 2007; van der Meer et al., 2011; Döhnel et al., 2012]. However, those previous studies cannot answer the question whether the pMPFC is involved in computing diverging beliefs or rather in resolving the resulting conflict by response selection, as could be argued based on parallel findings from stroop paradigms [see e.g., Venkatraman et al., 2009].

In the story phase of the present experiment, participants did not know if they would be asked for their own belief or the belief of the story protagonist. Thus at this time they did not have to resolve a response conflict and select one belief at the expense of the other. Observing pMPFC activity as early as in the story phase and not in the subsequent response phase suggests that this region is involved in the computation of diverging beliefs rather than in resolving a response conflict and selecting a response. This pattern of results is supported by recent findings on cognitive control in an adapted go/no-go task [Schulz et al., 2011]. To separate response selection from preparatory processes, a cue which indicated the location of the following go/no-go stimulus preceded the target. Surprisingly, activity in this region was not observed during the target phase related to response selection, but rather during the cue phase related to preparatory cognitive processes irrespective of the latter response. The results of the present study indicate that in belief reasoning, the primary role of the pMPFC appears not to be conflict resolution via the selection of one belief over another but rather to compute incongruent mental states, a process which may be essential for the decoupling mechanism.

Computing another's discrepant belief in the story phase was also associated with activity in the left and right temporoparietal cortex. This region was previously linked to reasoning about another's mental state [cf., Castelli et al., 2000; Schultz et al., 2003; Lombardo et al., 2010; Mar, 2011; van der Meer et al., 2011] and seems to play an important role in the processing of mental states that can be deduced from perceived actions [cf., Gobbini et al., 2007]. In the story phase of the present paradigm it was very important that the participants concentrated on the ball (did the ball jump back into the box *before* the boxes switched or did it jump in the other box *after* the position change of the boxes?). Based on the observed information, the participants had to update their own internal models of the world with respect to what they see and know and with respect to what the protagonist sees and knows. Whereas in the incongruent beliefs condition the action of the ball led to diverging mental models (the own belief in contrast to the other's false belief), the mental models in the congruent condition were the same for the participant and the protagonist. The activity of the bilateral temporoparietal cortex, associated with the computation of incongruent beliefs, may reflect the updating of mental models due to the changes in the situation.

This assumption is supported by the observed activity in the left inferior frontal gyrus (IFG) associated with the processing of diverging beliefs. Based on the character's visual access, his belief about the location of the ball had to be computed. For visual-perspective taking, it was recently shown that lateral frontal cortices are involved in selecting one's own as well as another's viewpoint [McCleery et al., 2011; Ramsey et al., 2013].

Effective Connectivity during the Computation of Incongruent Beliefs

To characterize the interplay of the pMPFC and bilateral temporoparietal cortex during the computation of incongruent beliefs in the story phase, we employed dynamic causal modelling combined with Bayesian model selection. In the resulting optimal model, the bilateral temporoparietal cortex (left and right pSTS/MTG) received driving input from experimental stimulation. When incongruent beliefs had to be computed, the strength of connections from the pMPFC to those temporoparietal regions increased. This is similar to recent findings by Hillebrandt et al. [2013], who reported influence of the superior dorsal MPFC to the left temporoparietal cortex in the presence of a social cue that invites for ToM reasoning. In the present paradigm, when incongruent beliefs had to be computed, activity of the pMPFC inhibited activity in temporoparietal cortices. Under the modulatory influence of incongruent beliefs, the connections pMPFC→left pSTS/MTG and pMPFC→right pSTS/MTG became negative, leading to reduced activity of the left and right pSTS/MTG. This fits well with prior observations that the pMPFC exerts goal-directed top-down influence [cf., Dosenbach et al., 2006; Schulz et al., 2011]. Also, a recent meta-analysis highlighted that the connections from this region to the temporoparietal cortex are dedicated to top-down-driven processing in social cognition [Bzdok et al., 2013]. Furthermore, Schulz et al. found in a psychophysiological interaction analysis that activity in this region was negatively correlated with activity in afferent regions of, for example, primary visual areas or the MTG. It was suggested that this relationship might reflect the inhibition of potential sensorimotor input. Activation of the pMPFC related to false belief reasoning was previously interpreted in terms of stimulus-independent processing [Sommer et al., 2007; Döhnel et al., 2012]: the other's false belief about the object's location has to be processed independently of the real location of the object (the own belief). In the current study the left and right pSTS/MTG may engage in the construction and adjustment of one's own and another's mental models based on relevant sensory input. When another's belief becomes false and no longer corresponds to the state of reality, stimulus-independent processing becomes necessary. This could be associated with activity in the

pMPFC, which inhibits the left and right pSTS/MTG in order to compute another's false belief, decoupled from reality and the own perception of the situation.

It is important to note that the computation of incongruent beliefs, in which the pMPFC exerts inhibitory influence, presumably entails at least two distinct cognitive operations: an encoding phase, in which the other's and one's own belief are generated, and a decoupling phase where those two diverging beliefs have to processed separately. The current design and temporal restrictions of fMRI do not allow distinguishing these processes. The present DCM analysis showed that the pMPFC inhibits the temporoparietal cortex, which received driving input from sensory stimulation, when beliefs were incongruent. In the light of this finding we propose that the pMPFC is not engaged in the initial encoding of beliefs, based on sensory input. It is rather active when a discrepancy between the other's and one's own belief is evident. Through the inhibition of stimulus-bound processing in temporoparietal regions, the pMPFC enables the decoupled processing of incongruent mental states. Further work is needed for a more fine-grained distinction between such two phases. As research on visual-perspective taking has shown, event-related potential paradigms can be a promising approach to address this issue [cf., McCleery et al., 2011].

These considerations regarding distinct subprocesses in the computation of incongruent beliefs give rise to a cautious interpretation of the detailed role of the left pSTS/MTG in this phase. We found that when it turned out that the other's and one's own belief were diverging, the negative intrinsic connection left pSTS/MTG→pMPFC became more negative. This modulation of the connection from the sensory-input-receiving pSTS/MTG to the pMPFC may reflect the role of the pSTS/MTG in signaling the discrepancy between beliefs to the pMPFC. Two sources in the literature fit this notion: In studies on visual-perspective taking and ToM reasoning (referring to a slightly more dorsal region of the temporoparietal cortex) the idea was brought up that the left temporoparietal cortex registers and indexes perspective differences [Aichhorn et al., 2006; Perner and Aichhorn, 2006; McCleery et al., 2011]. Furthermore two recent studies [Saygin et al., 2012; Hillebrandt et al., 2013] suggested that the left temporoparietal cortex (referring to an activation in the pSTS/MTG, close to the one observed in the current study) exerts bottom-up influence in the context of unexpected sensory information. A tentative interpretation of the current DCM findings is that an expected state could be "in the same situation another person is likely to perceive the world just as I do". If sensory information does not match this expectation ("Mine and the other person's perceptual access to the situation are diverging"), the left temporoparietal cortex signals this perspective difference to the pMPFC, which in turn initiates stimulus-independent processing through the inhibition of bilateral temporoparietal cortex and thus enables the decoupled processing of incongruent mental states. Future studies have to test this for a better understanding of the functional and temporal profile of the interplay of those regions during false belief reasoning.

Considering and Selecting Another's Belief (Question Phase)

In the question phase, subjects had to consider and select the previously encoded own or other's belief in order to give a correct response. Increased functional brain activity was observed contrasting the conditions in which the other's belief had to be considered (other-incongruent and other-congruent) with the conditions in which subjects were asked for their own belief (self-incongruent and self-congruent). Interestingly, the only brain region that showed increased functional activity in this contrast was a part of the right temporoparietal cortex, which is consistent with prior findings showing increased activity in this region for reasoning about another's in contrast to one's own mental state [Lombardo et al., 2010]. The differential activity of the temporoparietal cortex between processing another's and one's own belief contradicts assumptions by Legrand and Ruby [2009]. The authors proposed that activity in this region should not be affected by the type of belief, but rather by other task demands such as varying inferential processing or memory recall. However, those alternative factors were carefully controlled in the present study. Current findings tentatively suggest that the right temporoparietal cortex is particularly involved in processing another's mental state.

A conjunction analysis revealed that this subregion of the temporoparietal cortex showed also increased functional activity in the prior story phase (contrasting incongruent beliefs > congruent beliefs). Cabeza et al. [2012] highlighted that activity in this area is observed in multiple cognitive paradigms, related for example to attentional processes [Corbetta et al., 2000] or episodic memory retrieval [Wagner et al., 2005]. Cabeza et al. proposed a bottom-up attention hypothesis and argued that activity in this region related to attentional capture is not only elicited by stimuli in the environment, but also by internal information encoded in the memory. Viewing the current findings in the light of this hypothesis, we argue that in the story phase activity of the right temporoparietal cortex reflects the construction and adjustment of diverging mental models by encoding relevant environmental information. In the question phase the right temporoparietal cortex is involved in switching to and reconsidering the previously encoded mental model of the other in order to select the correct response.

CONCLUSION

We used conventional statistical parametric mapping and dynamic causal modelling in an adapted false belief paradigm to specify the neurocognitive components of ToM reasoning with respect to the two phases of (1)

computing incongruent beliefs and (2) considering and selecting either one's own or another's belief in order to give a response. Our findings suggest that the pMPFC and the bilateral temporoparietal cortex are crucially involved in the initial computation of diverging beliefs. The bilateral temporoparietal cortex may play a role in the construction and adjustment of mental models based on relevant sensory information in the environment. If another's belief is incongruent to the state of reality and one's own knowledge about a situation, the pMPFC might intervene in this process through the inhibition of activity in the bilateral temporal cortex and the suppression of stimulus-oriented processing. In this way it becomes possible to compute discrepant mental states and process another's false belief decoupled from one's own knowledge of the reality. In the subsequent selection of the relevant belief the right temporoparietal cortex might play a special role in switching to and reconsidering another's belief.

The current findings may help to generate explanations of continuing development of ToM reasoning during late childhood and adolescence long after the concept of false belief reasoning is established [Dumontheil et al., 2010; Vetter et al., 2012]. Besides substantial structural brain changes from late childhood to adulthood [e.g., Giedd, 2004; Mills et al., 2012] both the MPFC and the temporoparietal cortex show a developmental alteration of functional activity related to ToM reasoning [Moriguchi et al., 2007; Sommer et al., 2010; Meinhardt et al., 2011; Dumontheil et al., 2012; Gweon et al., 2012]. The question that has yet to be answered is whether these changes in functional activity reflect the maturation of the neural network underpinning ToM reasoning or whether they reflect a developmental change in cognitive processes constituting ToM reasoning per se [Blakemore, 2008; Burnett et al., 2011]. The investigation of qualitative and/or quantitative developmental changes in connectivity profiles between brain regions involved in ToM reasoning is a promising approach to address this question [cf., Burnett and Blakemore, 2009]. With regard to the current task, it might be the inhibitory influence of the pMPFC on the temporoparietal cortex, which is central to decoupling other's mental states, and still continues to develop during late childhood and adolescence.

ACKNOWLEDGMENTS

REFERENCES

Acs F, Greenlee MW (2008): Connectivity modulation of early visual processing areas during covert and overt tracking tasks. Neuroimage 41:380–388.

Aichhorn M, Perner J, Weiss B, Kronbichler M, Staffen W, Ladurner G (2009): Temporo-parietal junction activity in theory-of-mind tasks: Falseness, beliefs, or attention. J Cogn Neurosci 21:1179–1192.

Aichhorn M, Perner J, Kronbichler M, Staffen W, Ladurner G (2006): Do visual perspective tasks need theory of mind? NeuroImage 30:1059–1068.

Barbas H, Ghashghaei H, Dombrowski SM, Rempel-Clower NI (1999): Medial prefrontal cortices are unified by common connections with superior temporal cortices and distinguished by input from memory-related areas in the rhesus monkey. J Comp Neurol 410:343–367.

Barnes CL, Pandya DN (1992): Efferent cortical connections of multimodal cortex of the superior temporal sulcus in the rhesus monkey. J Comp Neurol 318:222–244.

Baron-Cohen S, Leslie AM, Frith U (1985): Does the autistic child have a ''theory of mind''? Cognition 21:37–46.

Beauchamp MS, Argall BD, Bodurka J, Duyn JH, Martin A (2004): Interhemispheric integration of visual processing during task-driven lateralization. Nat Neurosci 7:1190–1192.

Bird CM, Castelli F, Malik O, Frith U, Husain M (2004): The impact of extensive medial frontal lobe damage on 'Theory of Mind' and cognition. Brain 127:914–928.

Blakemore SJ (2008): The social brain in adolescence. Nat Rev Neurosci 9:267–277.

Botvinick MM, Cohen JD, Carter CS (2004): Conflict monitoring and anterior cingulate cortex: An update. Trends Cogn Sci 8: 539–546.

Burnett S, Blakemore SJ (2009): Functional connectivity during a social emotion task in adolescents and in adults. Eur J Neurosci 29:1294–1301.

Burnett S, Sebastian C, Kadosh KC, Blakemore SJ (2011): The social brain in adolescence: Evidence from functional magnetic resonance imaging and behavioural studies. Neurosci Biobehav R 35:1654–1664.

Bzdok D, Langner R, Schilbach L, Engemann DA, Laird AR, Fox PT, Eickhoff S (2013): Segregation of the human medial prefrontal cortex in social cognition. Front Hum Neurosci 7:232.

Cabeza R, Ciaramelli E, Moscovitch M (2012): Cognitive contributions of the ventral parietal cortex: An integrative theoretical account. Trends Cogn Sci 16:338–352.

Caspers S, Zilles K, Laird AR, Eickhoff SB (2010): ALE meta-analysis of action observation and imitation in the human brain. Neuroimage 50:1148–1167.

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

Corbetta M, Kincade JM, Ollinger JM, McAvoy MP, Shulman GL (2000): Voluntary orienting is dissociated from target detection in human posterior parietal cortex. Nature Neurosci 3:521–521.

DiQuattro NE, Geng JJ (2011): Contextual knowledge configures attentional control networks. J Neurosci 31:18026–18035.

Döhnel K, Schuwerk T, Meinhardt J, Sodian B, Hajak G, Sommer M (2012): Functional activity of the right temporo-parietal junction and of the medial prefrontal cortex associated with true and false belief reasoning. Neuroimage 60:1652–1661.

Dosenbach NUF, Visscher KM, Palmer ED, Miezin FM, Wenger KK, Kang HC, Burgund ED, Grimes AL, Schlaggar BL, Petersen SE (2006): A Core System for the Implementation of Task Sets. Neuron 50:799–812.

Dumontheil I, Apperly IA, Blakemore SJ (2010): Online usage of theory of mind continues to develop in late adolescence. Dev Sci 13:331–338.

Dumontheil I, Hillebrandt H, Apperly IA, Blakemore SJ (2012): Developmental differences in the control of action selection by social information. J Cogn Neurosci 24:2080–2095.

Friston KJ, Fletcher P, Josephs O, Holmes A, Rugg MD, Turner R (1998): Event-related fMRI: Characterizing differential responses. Neuroimage 7:30–40.

Friston KJ, Harrison L, Penny WD (2003): Dynamic causal modelling. Neuroimage 19:1273–1302.

Friston KJ, Holmes, AP, Worsley KJ, Poline JP, Frith CD, Frackowiak RSJ (1995): Statistical parametric maps in functional imaging: A general linear approach. Hum Brain Mapp 2: 189–210.

Friston KJ, Penny WD, Phillips C, Kiebel S, Hinton G, Ashburner J (2002): Classical and Bayesian inference in neuroimaging: Theory. Neuroimage 16:465–483.

Frith CD, Frith U (2012): Mechanisms of social cognition. Annu Rev Psychol 63:287–313.

Frith U, Frith CD (2003). Development and neurophysiology of mentalizing. Philos Trans R Soc Lond B Biol Sci 358:459–473.

Gallagher HL, Frith CD (2003): Functional imaging of 'theory of mind'. Trends Cogn Sci 7:77–83.

Garrido MI, Kilner JM, Kiebel SJ, Stephan KE, Friston KJ (2007): Dynamic causal modelling of evoked potentials: A reproducibility study. Neuroimage 36:571–580.

Giedd JN (2004): Structural magnetic resonance imaging of the adolscent brain. Ann NY Acad Sci 1021:77–85.

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

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

Gweon H, Dodell-Feder D, Bedny M, Saxe R (2012): Theory of mind performance in children correlates with functional specialization of a brain region for thinking about thoughts. Child Dev 83:1853–1856.

Hagmann P, Cammoun L, Gigandet X, Meuli R, Honey CJ, Wedeen VJ, Sporns O (2008): Mapping the structural core of human cerebral cortex. PLoS Biol 6:e159.

Hartwright CE, Apperly IA, Hansen PC (2012): Multiple roles for executive control in belief-desire reasoning: Distinct neural networks are recruited for self perspective inhibition and complexity of reasoning. Neuroimage 61:921–930.

Hillebrandt H, Dumontheil I, Blakemore S, Roiser JP (2013). Dynamic Causal Modelling of effective connectivity during perspective taking in a communicative task Neuroimage 76: 116–124.

Legrand D, Ruby P (2009): What is self-specific? Theoretical investigation and critical review of neuroimaging results. Psychol Rev 116:252–282.

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

Lombardo MV, Chakrabarti B, Bullmore ET, Wheelwright SJ, Sadek SA, Suckling J, MRC AIMS Consortium, Baron-Cohen S (2010): Shared neural circuits for mentalizing about the self and others. J Cogn Neurosci 22:1623–1635.

Ma L, Steinberg JL, Hasan KM, Narayana PA, Kramer LA, Moeller FG (2012): Working memory load modulation of parieto-frontal connections: Evidence from dynamic causal modeling. Hum Brain Mapp 33:1850–1867.

Mar RA (2011): The neural bases of social cognition and story comprehension. Ann Rev Psychol 62:103–134.

Mars RB, Sallet J, Schuffelgen U, Jbabdi S, Toni I, Rushworth MFS (2011): Connectivity-based subdivisions of the human right ''temporoparietal junction area'': Evidence for different areas participating in different cortical networks. Cereb Cortex 22: 1894–1903.

McCleery JP, Surtees ADR, Graham KA, Richards JE, Apperly IA (2011): The neural and cognitive time course of theory of mind. J Neurosci 31:12849–12854.

Meinhardt J, Sodian B, Thoermer C, Döhnel K, Sommer M (2011): True- and false-belief reasoning in children and adults: An event-related potential study of theory of mind. Dev Cogn Neurosci 1:67–76.

Mills KL, Lalonde F, Clasen LS, Giedd JN, Blakemore SJ (2012): Developmental changes in the structure of the social brain in late childhood and adolescence. Soc Cogn Affect Neurosci (Advance online publication).

Moriguchi Y, Ohnishi T, Mori T, Matsuda H, Komaki GEN (2007): Changes of brain activity in the neural substrates for theory of mind during childhood and adolescence. Psychiatry Clin Neurosci 61:355–363.

Neumann J, Lohmann G (2003): Bayesian second-level analysis of functional magnetic resonance images. Neuroimage 20:1346–1355.

Nichols T, Brett M, Andersson J, Wager T, Poline JP (2005): Valid conjunction inference with the minimum statistic. Neuroimage 25:653–600.

Penny WD, Stephan KE, Mechelli A, Friston KJ (2004): Comparing dynamic causal models. Neuroimage 22:1157–1172.

Penny WD, Stephan KE, Daunizeau J, Rosa MJ, Friston KJ, Schofield TM, Leff AP (2010): Comparing Families of Dynamic Causal Models. PLoS Comput Biol 6:e1000709.

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. Soc Neurosci 1:245–258.

Qureshi AW, Apperly IA, Samson D (2010): Executive function is necessary for perspective selection, not Level-1 visual perspective calculation: Evidence from a dual-task study of adults. Cognition 117:230–236.

Ramsey R, Hansen P, Apperly IA, Samson D (2013): Seeing it my way or your way: Frontoparietal brain areas sustain viewpoint-independent perspective selection processes. J Cogn Neurosci 25:670–684.

Ridderinkhof KR, Ullsperger M, Crone EA, Nieuwenhuis S (2004): The role of the medial frontal cortex in cognitive control. Science 306:443–447.

Rothmayr C, Sodian B, Hajak G, Döhnel K, Meinhardt J, Sommer M (2011): Common and distinct neural networks for false-belief reasoning and inhibitory control. Neuroimage 56:1705–1713.

Saxe R (2009): Theory of Mind (neural basis). In Banks W, editor. Encyclopedia of Consciousness. Claremont: Academic Press. pp 401–409.

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

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

Saygin AP, Chaminade T, Ishiguro H, Driver J, Frith CD (2012): The thing that should not be: Predictive coding and the uncanny valley in perceiving human and humanoid robot actions. Soc Cogn Affect Neurosci 7:413–422.

Schultz RT, Grelotti DJ, Klin A, Kleinman J, Van der Gaag C, Marois R, Skudlarski P (2003): The role of the fusiform face area in social cognition: Implications for the pathobiology of autism. Philos Trans R Soc Lond B Biol Sci 358:415–427.

Schulz KP, Bédard ACV, Czarnecki R, Fan J (2011): Preparatory activity and connectivity in dorsal anterior cingulate cortex for cognitive control. Neuroimage 57:242–250.

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

Sommer M, Meinhardt J, Eichenmüller K, Sodian B, Döhnel K, Hajak G (2010): Modulation of the cortical false belief network during development. Brain Res 1354:123–131.

Stephan KE, Harrison L, Kiebel S, David O, Penny WD, Friston KJ (2007): Dynamic causal models of neural system dynamics: Current state and future extensions. J Biosci 32:129–144.

Stephan KE, Penny WD, Daunizeau J, Moran RJ, Friston, KJ (2009): Bayesian model selection for group studies. Neuroimage 46:1004–1017.

Stephan KE, Penny WD, Moran JM, den Ouden HEM, Daunizeau J, Friston KJ (2010): Ten simple rules for dynamic causal modeling. Neuroimage 49:3099–3109.

Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M (2002): Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage 15:273–289.

van der Meer L, Groenewold NA, Nolen WA, Pijnenborg M, Aleman A (2011): Inhibit yourself and understand the other: Neural basis of distinct processes underlying Theory of Mind. Neuroimage 56:2364–2374.

Van Overwalle F (2009): Social cognition and the brain: A meta-analysis. Hum Brain Mapp 30:829–858.

Van Overwalle F, Baetens K (2009): Understanding others' actions and goals by mirror and mentalizing systems: A meta-analysis. Neuroimage 48:564–584.

Venkatraman V, Rosati AG, Taren AA, Huettel SA (2009): Resolving response, decision, and strategic control: Evidence for a functional topography in dorsomedial prefrontal cortex. J Neurosci 29:13158–13164.

Vetter NC, Leipold K, Kliegel M, Phillips LH, Altgassen M (2012): Ongoing development of social cognition in adolescence. Child Neuropsychol iFirst:1–15.

Visser M, Jefferies E, Embleton KV, Lambon R, Matthew A (2012): Both the middle temporal gyrus and the ventral anterior temporal area are crucial for multimodal semantic processing: Distortion-corrected fMRI Evidence for a double gradient of information convergence in the temporal lobes. J Cogn Neurosci 24:1766–1778.

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

Vossel S, Weidner R, Driver J, Friston KJ, Fink GR (2012): Deconstructing the architecture of dorsal and ventral attention systems with dynamic causal modeling. J Neurosci 32:10637–10648.

Wagner, A. D., Shannon, B. J., Kahn, I., Buckner, R. L. (2005): Parietal lobe contributions to episodic memory retrieval. Trends Cogn Sci 9:445–453.

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