Temporo-parietal Junction Activity in Theory-of-Mind Tasks: Falseness, Beliefs, or Attention

Markus Aichhorn[1], Josef Perner[1], Benjamin Weiss[1], Martin Kronbichler[1,2], Wolfgang Staffen[2], and Gunther Ladurner[2]

Abstract

■ By combining the false belief (FB) and photo (PH) vignettes to identify theory-of-mind areas with the false sign (FS) vignettes, we re-establish the functional asymmetry between the left and right temporo-parietal junction (TPJ). The right TPJ (TPJ-R) is specially sensitive to processing belief information, whereas the left TPJ (TPJ-L) is equally responsible for FBs as well as FSs. Measuring BOLD at two time points in each vignette, at the time the FB-inducing information (or lack of information) is presented and at the time the test question is processed, made clear that the FB is processed spontaneously as soon as the relevant information is presented and not on demand for answering the question in contrast to extant behavioral data. Finally, a fourth, true belief vignette (TB) required teleological reasoning, that is, prediction of a rational action without any doubts being raised about the adequacy of the actor's information about reality. Activation by this vignette supported claims that the TPJ-R is activated by TBs as well as FBs. ■

INTRODUCTION

The neuroimaging studies locating the processes for social cognition in the brain grow rapidly in number and provide a more detailed picture. A central ability consists in the attribution of mental states (mentalizing or theory of mind [ToM]). Early studies took a broad brush approach in order to capture all aspects of such mental attributions by comparing psychologically fairly rich stories (Fletcher et al., 1995) or cartoons (Gallagher et al., 2000) with descriptions or depictions of physical events. A first review of these early studies (Frith & Frith, 1999, 2003) found consistent activation across all these studies due to mentalizing in three bilateral areas: the medial prefrontal cortex (MPFC), in particular, the anterior cingulate cortex (ACC) and paracingulate area; adjacent areas of the parietal and temporal lobes (TPJ: temporo-parietal junction, and pSTS: posterior superior temporal sulcus); and the temporal poles (TP). Of these areas, only the one in the MPFC was thought of being specifically relevant for processing ToM. The other areas, it was thought, were needed because mental attribution typically involves animated, moving entities (parietal–temporal region) and social scripts (TP). Recent work has put this picture into flux. Functional ascriptions to areas are changing and subareas are being discovered that are responsible for different aspects of mental life.

Common sense concepts of the mind can be put into three classes (Hilgard, 1980): conative (wanting), cognitive (knowing/believing), and affective (emotions, feelings) aspects of mental life. A central purpose of mentalizing is to explain and predict intentional action, which can be defined as behavior that aims at changing given circumstances (what one knows to be the case) into desirable circumstances (goals: what one wants to be the case). A basic form of understanding intentional action consists in figuring out what needs to be done (rational action) in order to change existing circumstances (as they are known to us) into what is desirable (the goal of the action). This kind of explanation can be called goal–circumstance reasoning or *teleology* because it explains actions as behavior that aims at producing a particular goal. Behavior that suits this purpose (without unnecessary deviations) is also called a *rational action* (Csibra, Bíró, Koós, & Gergely, 2003; Blackburn, 1994; Cherniak, 1994). For instance, if the rock band plays in Munich and Anne wants to hear that band, then it is rational for her to drive to Munich.

Although the teleological approach will provide accurate predictions and explanations in most normal circumstances, there are a nonnegligible number of incidents where seemingly rational people appear to act irrationally, for example, Anne drives to Munich because she wants to hear the band but the band is not playing in Munich. Rationality can often be preserved by realizing that rational action does not depend on what is the case and what is desirable, but on what the actor thinks (believes) to be the case and deems to be desirable (wants), that is, rational action depends on how the world is represented. In other words, a seemingly irrational action

[1]University of Salzburg, Austria, [2]Paracelsus Private Medical University, Austria

can be understood as rational from the actor's perspective of how she represents the world, for example, Anne's pointless travel to Munich becomes understandable when we learn that she was misinformed about an alleged performance in Munich. This level of explanation has been dubbed a *representational theory of mind* as opposed to a situation theory (Perner, 1991) or *mentalism* as opposed to teleology (Csibra & Gergely, 1998). Goal–circumstance reasoning becomes belief–desire reasoning.

Emotions are reactions to the interplay of how our actions achieve or fail to achieve the desired changes, and they change the conative value of things. They thus serve a regulatory function in goal pursuit and rationality of action.

This short reminder of different approaches to explaining rational action (from teleology to a representational understanding of the mind) and the role of emotions helps us understand the more recent brain imaging findings on ToM. Evidence that the core processing for ToM cannot be located in the MPFC (Frith & Frith, 2003: paracingulate area) came from patients with lesions in relevant areas of the MPFC. Bird, Castelli, Malik, Frith, and Husain (2004) report on Patient G. T. with bilateral damage to the MPFC without detectable impairment on false belief tasks, tasks that assess understanding of belief–desire reasoning, that is, understanding rational action from a different perspective. Apperly, Samson, Chiavarino, and Humphreys (2004) found that patients with MPFC lesions did show impairment on false belief tasks but not as specifically as patients with lesions in the TPJ-L (left hemisphere). Studies on visual perspective taking (no reasoning about actions or their emotional consequences required) showed a glaring absence of any activation in the relevant areas of the MPFC (Aichhorn, Perner, Kronbichler, Staffen, & Ladurner, 2006; Vogeley et al., 2004; Zacks, Vettel, & Michelon, 2003). Consequently, this region is now thought of as centrally involved for processing emotional concerns with further subdivisions for different kinds of emotional involvement (Amodio & Frith, 2006).

Saxe and Kanwisher (2003) introduced a much more pointed contrast between a false belief task (FB) and structurally very similar photo vignettes (PH), namely, foreshortened text vignettes for adults of stories originally used for 3- to 5-year-old children (Zaitchik, 1990; Wimmer & Perner, 1983). In the FB task, an unwitnessed unexpected change of location leaves an observer with the FB that the transferred object is still in its original place. In the photo task, a change in location after a photo was made leaves the object in the photo in its original place. The contrast between these two vignettes also showed stronger activation in the temporo-parietal areas than in the MPFC. This and the lesion data from Apperly et al. (2004) and Samson, Apperly, Chiavarino, and Humphreys (2004) strengthened the view that belief–desire reasoning is associated with the TPJ more

than—as originally thought—with the MPFC. Thus, attention shifted to the TPJ as central for reasoning about action and appreciating perspective differences; in particular, in the right hemisphere (TPJ-R) because of repeated reports that Saxe's contrast between FB and PH was predominantly found on the right side (Mitchell, 2008; Young, Cushman, Hauser, & Saxe, 2007; Perner, Aichhorn, Kronbichler, Staffen, & Ladurner, 2006; Saxe & Powell, 2006; Saxe, Schulz, & Jiang, 2006; Saxe & Wexler, 2005).

The finding by Aichhorn et al. (2006) that the only "known" ToM area activated by their visual perspective tasks was in the TPJ-L led to the hypothesis that the TPJ-L is associated with perspective differences independently of rational (goal-oriented) action or emotions. Perner et al. (2006) pursued this hypothesis by adding a false sign task (FS: another foreshortened version of stories for children; Perner & Leekam, 2008; Parkin, 1994) to Saxe and Kanwisher's FB and PH task. The reasoning was that the FB task activates the TPJ-L more than the photo task does because FBs clearly raise a problem of differing perspectives—the believer thinks the world to be one way when the participant knows it to be different—whereas photos of the past do not create such a discrepancy. In particular, verbal descriptions of photos (as used by Saxe) do not make participants aware of the perspectivity of photos, that the photo shows the scene from a particular perspective. Although an FS, for instance, a direction sign, "castle," pointing the wrong way, shows the direction of the castle as being different from where it really is. However, unlike the FB task, it does not (explicitly) require or evoke belief–desire reasoning. Thus, the prediction was that the area of the TPJ-L that is activated more strongly by FB than by photo vignettes should also respond to FS vignettes. This prediction was confirmed, within a small cluster of only 36 voxels around (−48 −57 30, BA 39). There is also encouraging support from three patients reported in Samson et al. (2004) with lesions in the TPJ-L who had specific problems with FB tasks. These patients also were equally impaired on a truly false photo task (Apperly, Samson, Chiaverino, Bickerton, & Humphreys, 2007), in which the photo was supposed to show an object's current location. Nevertheless, the imaging results, pointing to the TPJ-L as an area dedicated to perspective differences, deserve replication due to their theoretical importance and given the size of the existing imaging evidence. One objective of the present study is to check whether these results can be replicated.

Perner et al. (2006) found a quite different effect of the new FS task in the TPJ-R, where FB activated more strongly than the PH task. In this region, FS activated less strongly than FB and not significantly more strongly than PH within a larger cluster of 192 voxels around (54 −57 27, BA 39). This result provides good support for Saxe and Kanwisher's (2003) original claim that the TPJ-R is specially attuned to processing belief information. This

claim, however, has recently come under attack by Mitchell (2008). A second objective of the present study was, therefore, to check whether this supportive result can be replicated.

Saxe and Kanwisher (2003) interpreted their findings of an activation difference between the FB and the PH task that the TPJ-R is responsible not just for processing information about FBs but about beliefs in general (i.e., also true beliefs). Their interpretation rested on indirect evidence, namely, on the fact that the activation pattern was similar whether the test question in the FB condition concerned the content of the FB, or the story fact. Sommer et al. (2007) contrasted directly an FB with a TB (true belief) picture story, but without including a PH control task, and reported stronger activation in the TPJ-R for FB than for TB in a large cluster of 311 voxels around (34 −54 24, BA 39). This result is, however, not incompatible with Saxe and Kanwisher's claim, which does not imply that FB stories might lead to more activation than TB stories. Their claim is that in the region where FB tasks activate more strongly than PH tasks, a TB task should also activate more strongly than a PH task.

In this context, it is wise to point out that TB tasks may differ in how strongly they evoke reasoning about beliefs. In contrast to FB tasks, where concern about the actor's belief is necessary for arriving at a correct solution, TB tasks can be solved correctly without any concern for the actor's beliefs. It can be dealt with by teleological reasoning, which sees action as a rational way to achieve a goal. TB tasks are, therefore, tasks where belief reasoning is optional. Tasks are, therefore, likely to differ in the probability that they lead to belief reasoning (i.e., highly and minimally suggestive TB tasks). This is best explained with the typical unexpected transfer examples used in, for instance, Sommer et al.'s (2007) belief tasks. Their task is a closer variant of the task by Wimmer and Perner (1983) than Saxe's vignettes. Participants see how a story character fails to witness (FB) or does witness (TB) an unexpected transfer of an object from a known place to a new place. In both cases, the unexpectedness of the transfer raises questions about beliefs (i.e., whether the story character did or did not witness it). Hence, this kind of TB task is highly conducive to thinking about beliefs because participants realize that the witnessing of the unexpected transfer prevented the character from developing an FB. This contrasts with cases where no unexpected events are happening. Then, plausibly, one does not or is less likely to get concerned about beliefs. One can predict what people will do simply on the basis of how the world is (one can rely by default that they believe it to be that way) and how the goal can be achieved given these facts. The stronger interpretation of Saxe and Kanwisher's claim that the TPJ-R is activated by TB tasks is that it is activated even in tasks where it is assumed to be true by default. The *third objective* of our study is to assess Saxe

and Kanwisher's claim by using such a minimally suggestive TB task to see whether it activates the TPJ-R more strongly than the PH task in that region where there is an activation difference between FB and PH.

Our *fourth objective* is to narrow down the time when the existence of a belief is inferred. This has interesting theoretical ramifications that have hardly been investigated. The question is whether reasoning about belief is triggered on-line (automatically) whenever relevant information is coming in, whether beliefs are inferred spontaneously when the incoming information is relevant for the purpose on hand, or whether beliefs are not computed on-line but only when a question is asked and it becomes apparent that beliefs have to be taken into consideration. One interesting theoretical consequence is that question-triggered belief processing poses a problem for modularity theories of belief–desire reasoning (e.g., Leslie, Friedman, & German, 2004) because modules are supposedly triggered automatically by relevant information.

Apperly, Riggs, Simpson, Chiavarino, and Samson (2006) provided behavioral data on this question. They used stories in which an FB about the location of an object was induced by failure to witness an unexpected transfer. Adults answered questions about the story protagonist's belief about the object's location as fast as questions about the object's location when they were instructed to pay attention to the protagonist's beliefs. However, without such instructions, answers about beliefs were given considerably more slowly than answers about the object's location. The authors concluded that participants did not automatically compute beliefs from relevant information but infer them spontaneously under certain conditions. In order to explore whether the time of belief inference also shows in all or just some of the relevant areas of the brain, we systematically restructured the vignettes used by Perner et al. (2006) so that the belief-inducing event is mentioned in the second sentence of the vignette and the question about the behavioral prediction requiring belief reasoning came as the third sentence. If belief-relevant information triggers belief reasoning automatically or beliefs are inferred spontaneously (we will not be able to distinguish these two possibilities), then relevant areas (e.g., TPJ-R) should be activated by the second sentence. Whereas, if belief reasoning is only triggered when needed for an answer, then relevant areas should not be activated before the question is asked.

In sum, using modifications of the story vignettes used by Perner et al. (2006), we pursue the following four objectives:

(1) Confirm and extend evidence that the TPJ-L is associated with tasks that create a difference in perspective, for example, FB as well as FS vignettes in comparison to photo vignettes and teleology tasks. Expected contrast: FB = FS > TB = PH.

(2) Confirm and extend evidence that the TPJ-R is associated specifically with attribution of mental states that involve a perspective difference (e.g., FBs) but not with tasks that involve a perspective difference but no mental state attribution (e.g., FSs). Expected contrast: FB > FS = PH.

(3) Test whether the area in the TPJ-R responsible for processing FB information (FB > PH) is also activated by action predictions where the question of perspective differences (informational insufficiencies) does not arise, for example, our version of the TB condition. Expected contrast for Saxe and Kanwisher's (2003) proposal: FB ≥ TB > PH; for replicating Sommer et al.'s results: FB > TB ≥ PH.

(4) Determine the time aspects of belief attributions. Are they automatically triggered or spontaneously inferred when informational complications are described, or only when belief–desire reasoning becomes necessary for answering a question? Expected contrasts: if automatically triggered, then FB > PH should activate in the TPJ-R when the second sentence of the vignette is read; if triggered by the question, then FB > PH contrast emerges only when the question is being asked.

METHODS

Participants

Twenty-one volunteers (8 women) participated in the fMRI study for payment. The median age was 24 years, ranging from 21 to 41 years. Before scanning, subjects were screened for neurological disorders and contradictions for MRI and had to give written informed consent. All participants were native German speakers, had normal vision and normal reading abilities.

Behavioral Procedure and Design

Functional images were obtained in one scanning session of 13 min, consisting of two recording runs. Participants had to read 32 vignettes on a screen in front of them, lasting 18 sec each. These trials were separated by short resting periods to be able to compute baseline comparisons. Each vignette consisted of two sentences and a question. The first sentence established a setting and introduced the protagonist, followed after 4 sec by the second sentence, which completed the story. Special attention was paid to the fact that in all vignettes the intentions of the involved protagonists could not be inferred before the second sentence was presented. Hence, the time interval in which the second sentence was presented and processed was considered critical for understanding the vignette and is, therefore, referred as "Time Point 1: Story."

After 7 sec, the question was presented with two answer choices. Within 5 sec participants had to choose the correct answer by selecting either the left or right option by pressing the corresponding button. To this time period we refer as "Time Point 2: Question." After an interstimulus interval of 2 sec, the next trial followed.

Our presented vignettes were modeled after the *Theory of Mind—Photo Contrast* vignettes by Saxe and collaborators (Young et al., 2007; Saxe & Powell, 2006; Saxe et al., 2006; Saxe & Wexler, 2005; Saxe & Kanwisher, 2003), Mitchell (2008), and in particular, the version created by Perner et al. (2006). Eight different scenes were created and each adapted for four different conditions: false belief reasoning (FB), false direction sign (FS), "false" photographs (PH), and true belief reasoning (TB). The sequence of conditions was randomized and the positions of correct and incorrect answer possibilities were randomly switched.

fMRI Data Acquisition and Analyses

MRI data were acquired on the Philips Gyroscan NT 1.5-Tesla scanner (Philips Medical System, Best, The Netherlands) of the Christian Doppler Clinic Salzburg. Blood oxygenation level dependent (BOLD) contrast images were measured using a gradient echo-planar imaging (EPI) sequence [TR = 2.3 sec, TE = 45 msec, FoV = 220 mm, flip angle (θ) = 90°, 64 × 64 matrix, slice thickness 4.5 mm (no gap), voxel size = (3.43 × 3.43 × 4.5) mm³]. Per volume 25 axial slices were acquired parallel to the AC–PC line, covering 112.5 mm of the z-axis. At each of the two runs, 163 EPI volumes were obtained. Prior to each run, six dummy images were acquired to allow transient signals to diminish.

After functional imaging, a structural MR image was acquired with an MP-RAGE sequence [TR = 8.6155 msec, TE = 4 msec, flip angle (θ) = 8°]. One hundred eighty-four (184) sagitally oriented slices (thickness = 1.2 mm) were acquired within a field of view of 240 mm. The FoV was reconstructed to a 256 × 256 matrix resulting in a voxel size of (1.2 × 0.9375 × 0.9375) mm³.

Image Analysis

Images were processed with SPM5 (Wellcome Department of Imaging Neuroscience, London, UK), implemented in a MATLAB 6.5 (Mathworks, Sherborn, MA) runtime environment. Images were slice-time and motion corrected by standard SPM5 algorithms. Subsequently, all functional images and high-resolution structural scans were coregistered and normalized to standard space (Montreal Neurological Institute [MNI], McGill, Montreal, Canada). The normalized images were resliced to isotropic Voxels of (3 × 3 × 3) mm³ and smoothed by a Gaussian kernel (8-mm FWHM).

For each participant, statistical analysis was performed using a general linear model, which modeled the differ-

ent components of one trial as boxcar function, convolved with a synthetic hemodynamic response function, its temporal and dispersion derivatives. Low-frequency noise was removed by a filter with a cutoff of 128 sec and serial correlations were taken into account using an autocorrelation model AR(1). Realignment parameters, session means, and a text-reading parameter were introduced as regressors of no interest. To be able to explore those regions which were already activated before the question is asked, we used different first-level designs to estimate the neuronal response at the two time points. We modeled the following components for the activations at *Time Point 1: Story*: one reading process spanning the first 4 sec followed by the four conditions lasting the seven critical seconds at the end of the story part. This time period was the second sentence of our stories and provided the information which differentiated our four conditions (FB, FS, TB, and PH). The answering process was modeled with the reaction times of each trial, and the missing seconds from the response to the end of each block. For the design at *Time Point 2: Question*, we used an overall reading phase of 11 sec preceding the question. The answering process was split up into the four conditions and again was followed by the same subsequent period at the end of each block (all onsets and durations for the trials were the same as in the design at *Time Point 1: Story*).

First-level contrasts were introduced into a repeated measures random effects analysis to allow for population inferences. To identify significantly activated clusters in the global analysis, a cluster-level threshold of $p_{corr.} < .05$ was used (corrected for multiple comparisons; height threshold of $p_{uncorr.} < .001$). Our additional analysis of voxel counts of the remaining comparisons within significant activated clusters in FB > PH contrast examined all voxels surviving a threshold of $p_{uncorr.} < .001$ in one of the remaining post hoc contrasts (see Table 1).

RESULTS

Reaction Times

The reaction times were analyzed with a repeated measures ANOVA and showed a main effect for conditions [$F(3, 60) = 50.49$, $\eta^2_{part.} = 0.716$, $p < .0001$; "false" photos: $M = 2577$ msec, $SE = 108$; false belief: $M = 2947$ msec, $SE = 142.7$; false direction signs: $M = 3221$ msec, $SE = 147.1$; true belief reasoning: $M = 3772$ msec, $SE = 141.7$]. Post hoc comparisons revealed significant differences between each pair of conditions (PH < FB: $SE = 90.1$, $p < .001$; FB < FS: $SE = 103.5$, $p < .01$; FS < TB: $SE = 96.8$, $p < .00001$). Differences with reaction time results in earlier studies (e.g., Perner et al., 2006) and possible interactions with neuronal responses are discussed later along with the imaging results. In 8% of the total 672 trials, subjects did not respond within the required time frame (no difference between the con-

ditions), and in additional 10% of the trials, the answer was wrong. Although subjects made less errors during the photo vignettes (9) than TB (24), FS (19), or FB (17), the conditions did not differ significantly [$\chi^2(3, n = 69) = 6.82$, $p = .8$].

Imaging Results

We focus our analysis by concentrating on those regions of the whole brain volume that show significant differences ($p_{corr.} < .05$ at cluster level) between FB and PH vignettes during *Time Point 1: Story* or at *Time Point 2: Question*. If this procedure revealed only an area for only one time point, then we tried to find a corresponding activation at lower thresholds for the other time point. Within these regions of interests, we analyze the effect of the other conditions by means of signal plots (Figure 2).

The seven regions in Table 1, which were activated more strongly by FB than by PH, have all been reported in previous studies using this contrast. A noticeable absence in Table 1 are regions in the anterior cingulate/paracingulate part of the medial prefrontal cortex (MPFC). This absence can be best explained by the assumption that activation in the MPFC depends on the emotional consequences of mentalizing and the fact that our vignettes induce very little concerns about emotional consequences. Activation in this area was minimal also in the study by Perner et al. and in the studies by Saxe and coworkers.

(1) **TPJ-R**: Lateral activations on the right side were in the TPJ-R consisting of 100 voxels. Table 1 also shows that this region was activated during question and story period. Figure 1 shows that the overlap of regions of significant activation differences during story and question was very large (black areas). The effect of all four conditions in each FB > PH region was investigated by two methods. Table 1 shows for each region those contrasts that were significant at voxel level ($p_{uncorr.} < .001$), and the number of voxels (k) for which such a significant difference was registered. Figure 2 backs up this picture by showing the signal plot for all four conditions against the resting condition (empty screen). In case of the TPJ-R, this analysis shows that at the time of the story only FB showed an activation difference against PH and the other vignettes. However, at the time the question was being asked, the difference between FB and TB vanished, and between FB and FS was strongly reduced. Instead, there were now a few voxels that showed a difference between FS and PH and between TB and PH. This change over time is also reflected in the corresponding signal plot in Figure 2$_{Question}$, which shows that the activation difference consists in a lesser deactivation from baseline by some conditions than by other conditions. The pattern of results suggests that the TPJ-R is activated at first only by FB, but when the question is being asked, then similar

Table 1. All Effects of Global Analysis at *Time Point 2: Question*, Corresponding Effects at *Time Point 1: Story*, and Voxel Counts of Remaining Contrasts within FB > PH Areas

| | | *Time Point 1: Story* | | | | | | *Time Point 2: Question* | | | | | |
| | | Cluster | | | Coordinates | | | Cluster | | | Coordinates | | |
Effect/Region	Contrast	k_E	$p_{corr.}$	Z	x	y	z	k_E	$p_{corr.}$	Z	x	y	z
Lateral–Right													
(1) TPJ-R	mid. temporal												
	FB > PH:	**66**	.014	4.83	57	−54	15	**100**	.002	4.61	54	−51	18
	FB > FS:	35						10					
	FB > TB:	36						−					
	FS > PH:	−						4					
	TB > PH:	−						5					
(2) aMTG	mid. temporal												
	FB > PH:	**190**	.000	5.52	51	−18	−15	**262**	.000	6.19	57	−15	−15
	FB > FS:	4						8					
	FB > TB:	8						6					
	FS > PH:	49						158					
	TB > PH:	67						119					
Lateral–Left													
(3) TPJ-L	angular/mid./sup. temporal												
	FB > PH:	**109**	.001	5.26	−57	−54	24	**200**	.000	5.40	−51	−57	21
	FB > FS:	−						1					
	FB > TB:	9						1					
	FS > PH:	18						93					
	TB > PH:	2						129					
(4) aMTG	mid./sup. temporal							**255**	.000	5.42	−57	−33	−3
	FB > PH:	86	.004	5.09	−54	−30	−6	169[a]					
	FB > FS:	1						−	temporal parts of #4:[a]				
	FB > TB:	2						−					
	FS > PH:	30						133[a]					
	TB > PH:	4						62[a]					
(5a) TP	mid./sup. temporal/pole												
	FB > PH:	84	.005	4.99	−51	−3	−27	local maximum at:[b]			−51	6	−15
	FB > FS:	1											
	FB > TB:	8											
	FS > PH:	32											
	TB > PH:	6											
(5b) IFG								frontal parts of #4:[a]					
	FB > PH:	20	.4	3.86	−45	30	−15	**86**[a]			−45	36	−12
	FS > PH:	3						59[a]					
	TB > PH:	2						30[a]					

Table 1. (*continued*)

| | | Time Point 1: Story | | | | | | Time Point 2: Question | | | | | |
| | | Cluster | | | Coordinates | | | Cluster | | | Coordinates | | |
Effect/Region	Contrast	k_E	$p_{corr.}$	Z	x	y	z	k_E	$p_{corr.}$	Z	x	y	z
Medial													
(6) PC	precuneus												
	FB > PH:	–	–	–		–		**56**	.027	4.13	6	−63	45
	FS > PH:							48					
	TB > PH:							11					
(7) SFG	medial sup. frontal												
	FB > PH:	59	.023	4.21	−9	30	60	**90**	.003	4.87	−15	36	54
	FS > PH:	11						41					
	TB > PH:	12						72					

The labels of centrally involved anatomical regions were obtained through the Anatomical Automatic Labeling Toolbox (according to the parcellation method described by Tzourio-Mazoyer et al., 2002). All computations referring to regions of interest were done through the SPM toolbox MarsBaR (Brett, Anton, Valabregue, & Poline, 2002).

[a]The cluster aMTG-L of the global analysis at the time point of the question showed widespread temporal activations, which extend into inferior frontal regions. We therefore report at Cluster #4, the statistic of the global analysis cluster at the aMTG-L. To get meaningful values for the voxel count, we split this main cluster by the means of predefined anatomical regions. Effect #4 reflects all significant voxels of the global cluster at the aMTG-L belonging to the temporal lobe and Effect #5b for those falling into the inferior frontal gyrus (IFG).

[b]For Cluster #5a at the time of the question, we had to create a meaningful region for the comparison of the signal plots (see Figure 2). We used for the temporal pole (TP) an inclusive mask incorporating all voxels of the global cluster aMTG-L (question) and those voxels which are already active during the story period. As this deviant procedure reveals no compatible values for Table 1, we only report the local maxima falling within TP.

processes occur in the TB and FS tasks. As a consequence, the contrast FB > TB vanishes and is replaced by TB > PH in the signal plot [$t(20) = 3.06, p = .006$]. To a lesser degree, the same happens with FS: The difference between FB and FS gets diminished [vanishes in the voxel count but remains in the signal plot: $t(20) = 2.82, p = .01$] at the time of the question, and it is amended by a significant FS > PH contrast in the signal plot [$t(20) = 2.64, p = .016$] and a few significant voxels.

These findings speak to two of our four theoretical objectives. *Objective 2*: We replicate the finding by Perner et al. (2006) that the TPJ-R is primarily activated by FB but not by FS, which supports the claim by Saxe and Kanwisher (2003) that a region within the TPJ-R is specialized for computing mental states. The fact that at the time of answering the question FS has less deactivation than PH (in the signal plot only) poses a slight problem. One explanation consistent with Saxe and Kanwisher could be that the question draws participants' attention to the fact of the sign potentially misleading people, that is, creating an FB in a potential audience.

Objective 3: The fact that TB activates the TPJ-R more than PH supports the other claim by Saxe and Kanwisher

Figure 1. Activations of *False Belief > Photographs* at *Time Point 2: Question* (gray) and overlapping areas with *Time Point 1: Story* (black). Beside the small diffuse activation at left temporal pole, there are no areas with greater or exclusive activation during the story.
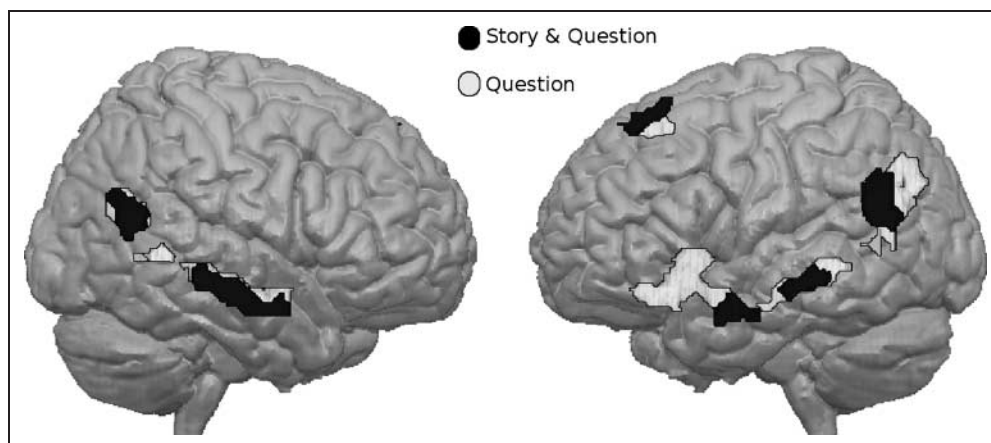
Figure 2. Parameter estimates and 95% confidence interval for all four conditions for *Time Point 1: Story* and *Time Point 2: Question* relative to resting phase. For error bars, normalized values with removed between-subjects variance according to Masson and Loftus (2003) were utilized. Bars indicate the 95% confidence interval. The signal plots incorporated all voxels within regions of interest defined by *False Belief > Photographs* contrast of Table 1. Due to the extant left temporal activations at *Time Point 2: Question*, we restricted the FB > PH contrast at the anterior part of the middle temporal gyrus (aMTG-L), the temporal pole (TP), and the inferior frontal gyrus (IFG) by means of inclusive masks to obtain meaningful regions (for details, see legend of Table 1).
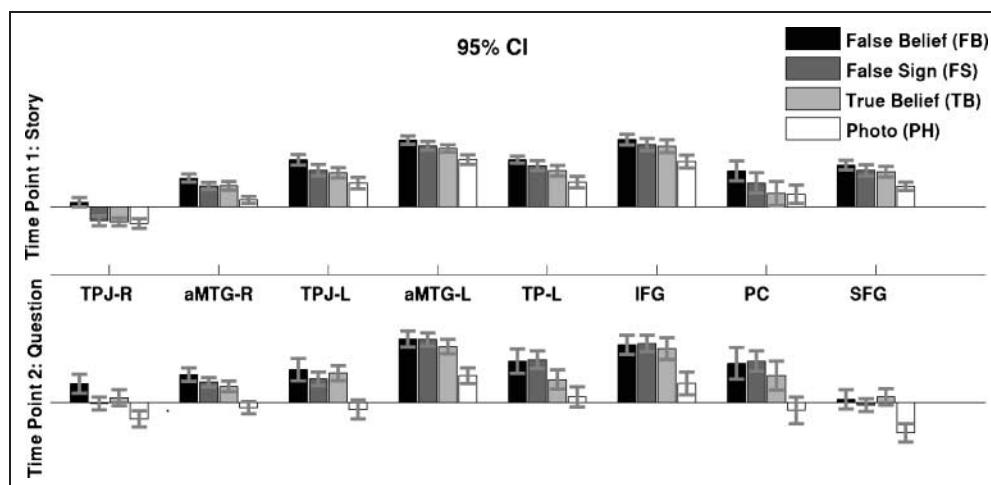
that also TB stories activate this region. This is the more remarkable, as we took care not to use a TB task, in which an unexpected change is witnessed by observers (e.g., Sommer et al., 2007), but a task in which no unexpected events occur. Nevertheless, one should point out that the activation could be due to the TB task occurring amidst other trials in which misinformation and FB have to be considered. So it might be that participants, when asked where the person in the TB vignettes will go to fulfil his goal (e.g., to get money), check out of caution whether there was, indeed, no misinformation about the relevant locations in play. However, this explanation cannot account for the fact that Sommer et al. (2007) found a difference between their FB and TB conditions, even though their TB condition was equally intermingled with FB stories. A reason for this discrepancy might lie in the timing of relevant information. Participants in Sommer's story may also have processed belief considerations at the time of the unexpected transfer but, because the TB condition made clear that the protagonist had all the necessary information, these processes had been laid to rest by the end of the story when the question was asked and BOLD was measured. In contrast, in our TB vignettes, only the question made apparent that the location information given in the story was relevant for predicting the protagonist's action. Hence, only then did participants check whether the protagonist had no information gaps or FBs. These considerations activated the TPJ-R as strongly as the FB or FS vignettes at the time of question.

In sum, the activation of the TPJ-R by TB vignettes speaks for the claim that belief considerations are processed in TB conditions or even when no informational problems are being made apparent (TB condition). However, we cannot exclude the possibility that the TPJ would have remained as inactive for TB as for PH if these conditions had been administered without constant reminders of the possibility of misinformation in the FS and FB vignettes. This possibility can only be tested in a new experiment with a quite different design of blocking TB trials off from FS and FB trials.

(2) **aMTG-R**: As frequently observed in previous studies (Perner et al., 2006; Saxe & Powell, 2006; Saxe et al., 2006; Saxe & Kanwisher, 2003), the FB > PH contrast shows activation in the more anterior part of the middle temporal gyrus (aMTG). Figure 1 shows that the activation at the time of the second sentence and the question overlaps strongly and forms a lengthy strand along the MTG. The voxel count in Table 1 and the signal plot in Figure 2 show a similar activation pattern to the TPJ-R. The activations due to FB, FS, and PH during the question help clarify the results by Perner et al. (2006), where FS was neither significantly different from FB nor from PH: FS is now clearly different from PH [$t(20) = 5.3$, $p < .0001$].

(3) **TPJ-L**: Figure 1 shows extensive activation of the TPJ-L. Although some early imaging studies of ToM reported stronger activation on the TPJ-L than on the TPJ-R region (Grèzes, Frith, & Passingham, 2004; Gallagher et al., 2000; Fletcher et al., 1995), practically all the recent studies using the FB > PH contrast reported stronger

activation differences in the right hemisphere (in case of Mitchell, 2008, there was no activation difference on left side at all). No ready explanation of this reversal in hemispheric asymmetry comes to mind.

The voxel count in Table 1 and the signal plots in Figure $2_{Question}$ show that the TPJ-L at the time of the question is activated about equally by FB, FS, and TB in relation to PH [all $ts(20) \geq 4.1$, $ps < .0005$], and that this process clearly starts at *Time Point 1: Story* for FB vignettes [$t(20) = 4.6$, $p = .0002$] and, to a lesser degree, for FS vignettes [$t(20) = 2.5$, $p = .05$]. These results speak to our *Objective 1*: The significant difference between FS and PH replicates the finding by Perner et al. (2006) that FS and FB (nonsignificantly different) both activate the TPJ-L in relation to PH during the question period. This finding is further underlined by the fact that, in contrast to the TPJ-R, FS differs from PH in the left hemisphere already when the misrepresentational aspect of the FSs is made apparent during *Time Point 1: Story*. An ANOVA computed with the extracted contrast estimates within volumes of interests at the TPJ-L and TPJ-R of FB and FS at *Time Point 1: Story* showed a significant Hemisphere × Condition interaction [$F(1, 20) = 5.18$, $p = .034$]. The lateralization is further underlined by a voxel count comparison based on spheres of 20 mm at left and right TPJ (LI-Toolbox of Wilke & Lidzba, 2007). This procedure revealed a clear right lateralization index of −0.8 already at a threshold of $t = 2.5$, which means that there were, at the given threshold, nine times more voxels at the right side compared to the left.

This pattern of findings implies that the TPJ at the left is not primarily responsible for processing information about mental states but seems to be equally responsible for processing misinformation and perhaps perspective differences in general (Aichhorn et al., 2006).

The fact that TB has a more than comparable effect on the TPJ-L does not fit this theory, under our original assumption that the TB vignettes would not induce any thoughts about differing perspectives. However, there is the possibility that when processing the question about where the story character will go, participants spontaneously started to make doubly sure that in this vignette the story character was—unlike the FB vignettes—not subjected to misinformation.

In this case, the TB vignettes would have elicited perspectival thoughts counter to what was intended. A better test of this possibility requires a study in which teleology (goal satisfaction) stories are tested in safe separation from stories involving misrepresentation and misinformation.

Also, the fact that at the time of story the PH condition did activate the TPJ-L above baseline poses some problem for this view compounded by the reaction time difference between conditions. The activation above baseline indicates that PH also requires processing of information in the TPJ-L. Insofar as PH does not involve a perspective contrast, this speaks against a specific responsibility of the TPJ-L for processing perspective contrasts. Alternatively, one can suspect that, although PH does not require processing of a perspective conflict, the simple mention of pictures brings up thoughts about perspectives. PH is not an ideal control for FB and FS when it comes to investigating perspective processes. This also fits the fact that FB and FS activated more strongly than PH and that, by the time of the question, activation in PH had fallen back to baseline. This pattern of activation comes about because in FB and FS, the perspective contrast remained relevant to the end, whereas in PH it has become apparent that the potential perspectival nature of photos is irrelevant.

(4) **aMTG-L**: The FB > PH covers 255 voxel along the MTG, the temporal pole, and inferior frontal gyrus (IFG). The activation was slightly more widespread than at the right hemisphere, therefore, we had to use predefined anatomical regions of interests to split this cluster into comparable regions for Table 1 (for details, see legend of Table 1). The activation pattern within the temporal parts of the cluster are similar to those in the aMTG-R: FB, FS, and TB activated significantly more strongly than PH and this process started already when processing the second sentence [all $ts(20) \geq 3.5$, $ps \leq .005$].

(5a) **TP**: Table 1 shows a large cluster of 84 voxels in the middle and superior temporal gyrus and the temporal pole during processing of the second sentence. The activation pattern in that region (Figure 2_{Story}) is similar to its adjacent regions. (FB ≥ FS ≥ TB > PH). At the time the question was asked, this cluster became so large that we compared in the signal plot of the temporal pole only those voxels which were already active during the story period. Nevertheless, the activation pattern changed only slightly, but FS is now significantly stronger than TB and PH [all ts (20) > 3.2, ps < .005].

This suggests that the description of TP containing social scripts (Frith & Frith, 2003) needs to be amended. For instance, it may get especially activated if the social script promises to be interesting. The FB vignettes make that promise already in the second sentence. The FS vignettes lag behind, but FS puts the potential complications ensuing from misinformation into the air and TB sets up a purpose and participants have to predict how best to fulfil that purpose. In contrast, the PH vignettes simply describe the taking of a photo and an environmental change. No further purpose, no complications of misinformation provided at any time point.

(5b) **IFG**: A substantial part (86 voxel) of the cluster with its peak voxel at the MTG-L at the time of the question extents into the IFG pars triangularis and orbitalis. Although unusual, we decided to split this region off the main cluster and analyze it separately, in particular, because this region becomes even more dominant at the time of the question. The signal plot shows a well-known pattern at the reported regions of the left hemisphere—FB, FS, and TB more or less

equally activated but significantly different from PH [all *t*s (20) > 3.5, *p*s < .002].

This region, a portion of Broca's area, is relevant for semantic processing of sentences and, therefore, the activation differences might be due to differential demands on language processing. However, no obvious difference in linguistic processing demands of these tasks compared to PH is apparent. Although none of the studies that contrasted FB with PH reported any activation differences in this or a nearby area, other ToM imaging studies do report activity comparing pretend versus real actions (German, Niehaus, Roarty, Giesbrecht, & Miller, 2004), mentalizing versus nonmentalizing (Mitchell, Banaji, & Macrae, 2005), third-person versus first-person perspective (Ruby & Decety, 2003), or false versus correct expectations (Grèzes et al., 2004). Several studies report activations in the adjacent orbital region (Walter et al., 2004; Berthoz, Armony, Blair, & Dolan, 2002) or in the temporal pole (Gallagher et al., 2000; Fletcher et al., 1995). Perhaps our results reflect the lower semantic processing demands of our photograph task as indicated by the shorter reaction times compared to photo stories in other studies and by the unexpected strong global activation level of the other conditions against PH. Thus, we modeled all trial-by-trial reaction time influences, but the anterior temporal and inferior frontal activations could not be explained as a simple function of the reaction times.

(6) **PC**: In line with most previous studies using the FB > PH contrast (Mitchell, 2008; Young et al., 2007; Perner et al., 2006; Saxe & Powell, 2006; Saxe et al., 2006; Saxe & Wexler, 2005; Saxe & Kanwisher, 2003), differential activation in the precuneus was found at 6 −63 45, which is close to the peak voxels reported previously within the region of: $x \in [-15, +15], y \in [-54, -66], z \in [20, 40]$. The pattern helps clarify the indeterminate result by Perner et al. (2006) because this time also FS clearly activated more strongly than PH, which showed the usual deactivation of the precuneus as part of the default activation system (e.g., Gusnard & Raichle, 2001; Raichle et al., 2001). Table 1 also shows that this is the only region where, at *Time Point 1*, no significant voxel could be found within any of the possible contrasts, but Figure 2 indicates that FB tends to activate more strongly than TB and PH [all *t*s (20) > 2.1, *p*s < .05].

(7) **SFG**: In the dorsal medial part of the superior frontal gyrus, we find activation of FB, FS, and TB in contrast to PH (peak voxel at −15 36 54). An FB > PH contrast in this region has been reported only by Mitchell (2008) (peak voxel: 24 27 57, *k* = 35, SFG) and in a nearby region by Saxe and Powell (2006) (peak voxel: 0 45 48, *k* = 11, in the dMPFC).

DISCUSSION

This study explored the function of those cerebral areas that show specific activation for false belief scenarios over photo scenarios. This contrast has recently been frequently used to narrow down relevant ToM areas starting with the study by Saxe and Kanwisher (2003). With the exception of the ACC/paracingulate area in the MPFC, probably due to the minimal emotional content of our vignettes (see Perner et al., 2006), all areas previously reported for this contrast came out again very strongly (TPJ-R, TPJ-L, aMTG/STS-R, TP-R, PC). One difference was that, in this study, the FB–PH contrast tended to be more left-lateralized in the TPJ than the usual right-hemisphere bias. As a consequence, also the aMTG-L, TP-L, and IFG-L showed strong activation differences that have not been reported in previous studies. The prime motivation for this study, however, was the pursuit of several objectives:

Objective 1 was to replicate the finding reported by Perner et al. (2006) that the TPJ-L is activated not only by FB but also by FS vignettes in relation to PH vignettes and to establish their interpretation more firmly, that the TPJ-L may be associated with computing perspective differences (also activated by visual perspective tasks, e.g., Aichhorn et al., 2006; Vogeley et al., 2004; Zacks et al., 2003). The replication succeeded extremely well: At the time the question was processed, the TPJ-L was equally active for FB and FS in contrast to PH (Figure 2). Moreover, the same pattern was also seen in aMTG-L. However, the interpretation of this finding by Perner et al. could not be clearly confirmed because of the activations generated by the new TB vignettes. In this condition, no perspective difference should be induced. It is a straight teleological problem of goal pursuit. Yet activation of the TPJ-L was as strong as for the FB and FS vignettes.

Let us consider two possible explanations. One of them stays close to redescribing the data: The TPJ-L is activated by concerns about rational actions (teleology) whether they concur with a perspective difference or not (hence, TB activates), but it is also activated by perspective differences even when no rational action is at stake (FS). Alternatively, we can stick to the original hypothesis but assume that TB, counter to our intentions, did trigger concerns about perspective differences. This occurred because the vignettes are within a context of informational insufficiencies portrayed in FB and FS vignettes. Hence, although there is nothing untoward reported in TB vignettes, participants still wanted to check whether they had not missed any informational confusions before answering the question. This could explain why responses to TB vignettes were slower than responses in the three other conditions. This possibility is also supported by the fact that TB > PH has a peak voxel in the TPJ-L above $z = 21$, when almost all imaging studies that investigated attribution of intentional actions without perspective differences show activation peaks below that line (Perner & Leekam, 2008; Gobbini, Koralek, Bryan, Montgomery, & Haxby, 2007; Aichhorn et al., 2006). To resolve the issue, the following design is

needed: Participants are first tested with TB and PH vignettes only and are then given the tracer contrast of FB > PH. This should prevent unintended concerns about perspective differences in the TB condition.

Objective 2 was to provide support for the claim by Saxe and Kanwisher (2003) that the TPJ-R is especially associated with computing mental states, such as FBs, by replicating the finding that FB vignettes activate more strongly than FS and PH vignettes in the TPJ-R (Perner et al., 2006). This aspect was clearly replicated (Figure 2_{Story}), although there was some indication that activation by FS was somewhat stronger than by PH. Moreover, in the more or less contiguous cluster of activations in the aMTG-R (Figure 2_{Question}), FB and FS clearly activated more strongly than PH. Nevertheless, the data do suggest that there may be some region within the TPJ-R responsible for FB over and above FS and PH.

Mitchell (2008) challenged Saxe and Kanwisher's (2003) interpretation of their finding because basic attentional processes, in particular, attention controlled by external stimuli, also activate in this region (Corbetta & Shulman, 2002). Mitchell showed that activations by one such attention task, a version of Posner, Walker, Friedrich, and Rafal's (1984) invalid cueing versus valid cueing task, overlaps substantially with activations by the FB > PH contrast. We concur with Mitchell that this indicates that there is some processing component common to FB tests and attention tasks not shared by the photo task. Mitchell favors a basic attentional process but is not very explicit on which process would apply to FB but not to PH. One possible common element could be the aspect of misinformation. This pertains to FBs and to invalid cues, but not to the photo task. However, Mitchell argues against this explanation by citing evidence that a task manipulating attention by flankers, not misleading cues, activates in the same region (Serences et al., 2005). This objection to misinformation as the common element is supported by our finding that FB vignettes activate in this region more than FS, which has misinformation as its feature.

This still leaves us without explanation of why the FB > PH (and FS) contrast activates the TPJ-R in the same region as the invalid > valid cue paradigm. Pursuing the solution of finding a common element, we suggest the following: The region in question activates when a *wrong action or action tendency is induced by misinformation*. This applies to FB but not to PH or FS (because no action is described). It also applies to Posner's invalid > valid cue contrast as invalid cues trigger wrong responses or at least wrong response tendencies, leading to hesitation (slower reaction times). The same can be claimed for the distracting flanker task used by Serences et al. (2005). Although, unlike invalid cues (arrow points to left when target appears on right), flankers (same color as color of target) are not explicitly misinforming that a target is present, but they willy-nilly create the tendency to respond with ''target present'' increasing the error rate. Thus, the distracting flankers contain implicit (unconscious) misinformation insofar as our action system is distracted by them. As a result, we have a case of an erroneous action tendency triggered by implicit misinformation. In contrast to this strategy of solving the problem by finding a common element, Saxe (personal communication) seeks a solution in higher imaging precision: FB > PH and the invalid > valid contrasts activate proximal but distinct subregions of the TPJ-R.

Objective 3 was to investigate Saxe and Kanwisher's claim that the TPJ-R not only activates for FBs but also for TBs, which was based on indirect reasoning. Sommer et al., who did use a TB condition reported that FB activated the TPJ-R (or some subregion) more strongly than TB. This still leaves the possibility that TB might activate the TPJ-R more strongly than PH. Our version of the TB vignettes were to test this possibility in an even stricter sense, because unlike Sommer et al.'s version, our's was not supposed to even raise the specter of misinformation. Figure 2_{Story} shows that, indeed, TB activated the TPJ-R more strongly than PH, as predicted by Saxe and Kanwisher (2003).

Our support for Saxe and Kanwisher's claim remains, however, tentative for the reasons discussed in relation to Objective 1 above. That is, reaction times to TB vignettes were slowest and TB > PH peaked in the TPJ-L above $z > 20$, when other ToM tasks that do not raise perspective differences tend to have peaks below that line. Hence, we surmised that because TB vignettes were presented within the context of FB and FS tasks, participants started checking the informational conditions in the TB vignettes, which raised the specter of misinformation and perspective differences. In that case, the conclusion that all TB tasks activate the TPJ-R would be misleading. In our case, TB might have activated the TPJ only because, counter to our intention, it was treated as a potential misinformation task by our participants. Another reason for why reaction times in TB were longest and caused activation in the TPJ could be that Sentence 2 in the TB vignettes is a pragmatically strange intrusion into the flow of the stories because it contains a semantically confusable but irrelevant distractor fact.

Objective 4 was to address the question of whether ToM considerations are triggered automatically or spontaneously inferred when relevant information occurs in the input or whether relevant inferences are only made when a question needs to be answered. Modularity theories of ToM (e.g., Friedman & Leslie, 2004) assume that modular computations are automatic. This is also considered an important feature by theorists who assume that ToM computations are necessary for linguistic, pragmatic reasons, that is, for keeping us up-to-date on our conversation partners' informational needs in order to adjust our speech acts accordingly (e.g., Sperber & Wilson, 2002). However, recently, Apperly et al. (2006) provided behavioral evidence that belief processing is not automatic. Participants needed longer time to answer

questions about a person's FB about an object's location than to answer the question about the object's location (equated for grammatical complexity), unless participants were instructed to pay particular attention to the person's belief. Our data quite unanimously show that beliefs are inferred spontaneously in the second sentence of the FB vignettes. Figure 2 makes clear that in every single area that shows the specific FB > PH activation difference at question time, that difference already shows up when the story is being processed. The evidence for early processing of FB vignettes in *all* the relevant regions makes it unlikely that early processing is due to merely registering the misinformation without also computing the resulting belief.

This finding does not contradict the claim made by Apperly et al. (2006) based on behavioral data that beliefs are not computed automatically. Our data do show that under our presentation conditions, beliefs are inferred spontaneously. A possible reason for this early processing of belief information could be the potential context effect we mentioned in connection with the TPJ activations caused by the TB vignettes (see discussion of Objective 1 above). If the context of FB and FS vignettes, according to assumption, is able to make participants suspect informational problems where there are none (i.e., our TB vignettes), then it is plausible to assume that this context of repeated exposure to FBs and misleading sign vignettes makes them anticipate and pay attention to informational conditions and their consequences where they would not do so spontaneously without this context. To assess this possibility, another study is required with a different design, in which FB vignettes are sporadically interspersed with TB and PH vignettes (and other informationally innocent scenarios). In contrast, in the behavioral study by Apperly et al., such context effects were reduced because diverse filler scenarios were used and judgments about range of other facts and beliefs were asked, which made it more difficult for participants to identify the relevant belief to encode. Moreover, we used verbal descriptions of scenes while Apperly et al. presented the information pictorially without verbal explanation. Plausibly, when just observing unwitnessed location changes, participants do not infer the resulting FB on-line, but they do so when this lack of information is described verbally because verbal information, assumed to be relevant, is understood as highlighting the important aspects of an event for later purposes.

Conclusion

Our results confirm the hemispheric functional asymmetry of the TPJ. The TPJ-R is specially sensitive to processing information about beliefs, false as well as true ones, as Saxe and Kanwisher (2003) suggested, whereas the TPJ-L is also concerned with nonmental cases of per-

spective differences (Aichhorn et al., 2006) and of misinformation (e.g., FSs; Perner et al., 2006). All regions that are sensitive to FB in contrast to photo vignettes showed early activation when the information about the formation of a belief became available (second sentence) and not on demand when the test question required it. The reason for such early, spontaneous belief inferences, in contrast to Apperly et al.'s behavioral findings, may lie in a potential ''context effect'' of vignettes that depict informational imperfections (FB and FS) and alerting participants to the possibility of such imperfection in later vignettes. This context effect could therefore create the erroneous impression that people always make early belief inferences spontaneously even in informationally innocent scenarios (our TB version), when in fact they would not do so in normal circumstances. Future research needs to look into this methodological issue, which is one of those research scenarios in which a separate localizer experiment provides the appropriate design (Friston, Rotshtein, Geng, Sterzer, & Henson, 2006) to break the unwanted context effect.

APPENDIX: EXAMPLE STIMULI

Example Vignettes of the Four Conditions as English Translation (Left) of the Original German Text (Right)

False Belief

Julia sees the ice cream van go to the lake.	Julia sieht den Eiswagen Richtung See fahren
She doesn't see that the van turns off to the town hall.	Sie sieht nicht, wie der Eiswagen zum Rathaus abbiegt
Therefore, Julia will look for the ice cream van at the …	Julia sucht daher den Eiswagen beim
Lake/Town Hall	See/Rathaus

False Sign

The ice cream vendor's sign points to the lake.	Das Eisverkäuferschild zeigt zum See
The ice cream van goes to the town hall without changing the sign.	Der Eiswagen fährt zum Rathaus ohne das Schild zu ändern
According to the sign post the ice cream van is at the …	Laut Schild ist der Eiswagen jetzt beim
Lake/Town Hall	See/Rathaus

False Photo

Julia takes a picture of the ice-van in front of the pond	Julia macht ein Photo vom Eiswagen vorm Teich

APPENDIX (*continued*)

The ice cream van changes to the market place; the picture gets developed	Der Eiswagen fährt zum Stadtplatz weiter, das Photo wird entwickelt
On the picture the ice-van is at the . . .	Auf dem Photo ist der Eiswagen vorm
Pond/Market place	Teich/Stadtplatz
True Belief	
Julia works at the ice cream van in front of the pond	Julia arbeitet vormittags beim Eiswagen am Teich
On her way home she passes the fruit merchant in the town square.	Am Heimweg geht sie beim Obsthändler am Stadtplatz vorbei
At noon she wants ice-cream, she walks back to the . . .	Mittags will Julia ein Eis, sie geht zum
Pond/Town square	Teich/Stadtplatz

Acknowledgments

REFERENCES

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

Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: The medial frontal cortex and social cognition. *Nature Reviews Neuroscience, 7,* 268–277.

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

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

Apperly, I. A., Samson, D., Chiavarino, C., & Humphreys, G. W. (2004). Frontal and temporo-parietal lobe contributions to theory of mind: Neuropsychological evidence from a false-belief task with reduced language and executive demands. *Journal of Cognitive Neuroscience, 16,* 1773–1784.

Berthoz, S., Armony, J. L., Blair, R. J. R., & Dolan, R. J. (2002). An fMRI study of intentional and unintentional (embarrassing) violations of social norms. *Brain, 125,* 1696–1708.

Bird, C. M., 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.

Blackburn, S. (1994). *The Oxford dictionary of philosophy.* Oxford: Oxford University Press.

Brett, M., Anton, J., Valabregue, R., & Poline, J. (2002). Region of interest analysis using an SPM toolbox. Presented at the 8th International Conference on Functional Mapping of the Human Brain, June 2–6, Sendai, Japan. *Neuroimage, 16,* 497.

Cherniak, C. (1994). Rationality. In S. Guttenplan (Ed.), *A companion to the philosophy of mind* (pp. 526–531). Oxford: Blackwell.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience, 3,* 201–215.

Csibra, G., Bíró, S., Koós, O., & Gergely, G. (2003). One-year-old infants use teleological representations of actions productively. *Cognitive Science, 27,* 111–133.

Csibra, G., & Gergely, G. (1998). The teleological origins of mentalistic action explanations: A developmental hypothesis. *Developmental Science, 1,* 255–259.

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,* 109–128.

Friedman, O., & Leslie, A. M. (2004). Mechanisms of belief–desire reasoning. Inhibition and bias. *Psychological Science, 15,* 547–552.

Friston, K. J., Rotshtein, P., Geng, J. J., Sterzer, P., & Henson, R. N. (2006). A critique of functional localisers. *Neuroimage, 30,* 1077–1087.

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

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

Gallagher, H., Happé, F., Brunswick, N., Fletcher, P., Frith, U., & Frith, C. (2000). Reading the mind in cartoons and stories: An fMRI study of "theory of mind" in verbal and nonverbal tasks. *Neuropsychologia, 38,* 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,* 1805–1817.

Gobbini, M. I., Koralek, A. C., Bryan, R. E., Montgomery, K. J., & Haxby, J. V. (2007). Two takes on the social brain: A comparison of theory of mind tasks. *Journal of Cognitive Neuroscience, 19,* 1803–1814.

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

Gusnard, D., & Raichle, M. (2001). Searching for a baseline: Functional imaging and the resting human brain. *Nature Reviews Neuroscience, 2,* 685–694.

Hilgard, E. R. (1980). The trilogy of mind: Cognition, affection, and conation. *Journal of the History of the Behavioral Sciences, 16,* 107–117.

Leslie, A. M., Friedman, O., & German, T. P. (2004). Core mechanisms in "theory of mind". *Trends in Cognitive Sciences, 8,* 528–533.

Masson, M., & Loftus, G. (2003). Using confidence intervals for graphically based data interpretation. *Canadian Journal of Experimental Psychology, 57,* 203–220.

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

Mitchell, J. P., Banaji, M. R., & Macrae, C. N. (2005). The link between social cognition and self-referential thought

in the medial prefrontal cortex. *Journal of Cognitive Neuroscience, 17,* 1306–1315.

Parkin, L. (1994). *Children's understanding of misrepresentation*. PhD thesis, University of Sussex.

Perner, J. (1991). *Understanding the representational mind. Bradford book*. 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,* 245–258.

Perner, J., & Leekam, S. (2008). The curious incident of the photo that was accused of being false: Issues of domain specificity in development, autism, and brain imaging. *Quarterly Journal of Experimental Psychology, 61,* 76–89.

Posner, M. I., Walker, J. A., Friedrich, F. J., & Rafal, R. D. (1984). Effects of parietal injury on covert orienting of attention. *Journal of Neuroscience, 4,* 1863–1874.

Raichle, M., MacLeod, A., Snyder, A., Powers, W., Gusnard, D., & Shulman, G. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences, U.S.A., 98,* 676–682.

Ruby, P., & Decety, J. (2003). What you believe versus what you think they believe: A neuroimaging study of conceptual perspective-taking. *European Journal of Neuroscience, 17,* 2475–2480.

Samson, D., Apperly, I. A., Chiavarino, C., & Humphreys, G. W. (2004). Left temporoparietal junction is necessary for representing someone else's belief. *Nature Neuroscience, 7,* 499–500.

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

Saxe, R., Schulz, L. E., & Jiang, Y. V. (2006). Reading minds versus following rules: Dissociating theory of mind and executive control in the brain. *Social Neuroscience, 1,* 284–298.

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

Serences, J. T., Shomstein, S., Leber, A. B., Golay, X., Egeth, H. E., & Yantis, S. (2005). Coordination of voluntary and stimulus-driven attentional control in human cortex. *Psychological Science, 16,* 114–122.

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.

Sperber, D., & Wilson, D. (2002). Pragmatics, modularity and mind-reading. *Mind & Language, 17,* 3–23.

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

Vogeley, K., May, M., Ritzl, A., Falkai, P., Zilles, K., & Fink, G. (2004). Neural correlates of first-person perspective as one constituent of human self-consciousness. *Journal of Cognitive Neuroscience, 16,* 817–827.

Walter, H., Adenzato, M., Ciaramidaro, A., Enrici, I., Pia, L., & Bara, B. G. (2004). Understanding intentions in social interaction: The role of the anterior paracingulate cortex. *Journal of Cognitive Neuroscience, 16,* 1854–1863.

Wilke, M., & Lidzba, K. (2007). LI-tool: A new toolbox to assess lateralization in functional MR-data. *Journal of Neurosciences Methods, 163,* 128–136.

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.

Young, L., Cushman, F., Hauser, M., & Saxe, R. (2007). The neural basis of the interaction between theory of mind and moral judgment. *Proceedings of the National Academy of Sciences, U.S.A., 104,* 8235–8240.

Zacks, J. M., Vettel, J. M., & Michelon, P. (2003). Imagined viewer and object rotations dissociated with event-related FMRI. *Journal of Cognitive Neuroscience, 15,* 1002–1018.

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