Do visual perspective tasks need theory of mind?

Markus Aichhorn, Josef Perner, Martin Kronbichler, Wolfgang Staffen, and Gunther Ladurner

Center of Neurocognitive Research, University of Salzburg, Austria
Department of Psychology, University of Salzburg, Austria
Department of Neurology, Christian Doppler Clinic, Paracelsus Privat Medical University, Salzburg, Austria

Received 20 July 2005; revised 22 September 2005; accepted 20 October 2005
Available online 6 December 2005

The concept of perspective has become a frequent object of investigation in neural imaging studies. Its popularity is grounded in the central role it plays in psychology and philosophy. With Nagel’s (1974) famous paper, it has become the hallmark of consciousness within philosophy (Block, 1995; Chalmers, 1996) and, more recently, neuroimaging of consciousness (Ruby and Decety, 2003, 2004; Vogele et al., 2001, 2004). These studies use imaging techniques to localize perspective relevant processes in the brain under the title of perspective and another group of studies under the rubric of “theory of mind” contrasting a wrong view of the world (false belief) with reality (Fletcher et al., 1995; Gallagher et al., 2000; Goel et al., 1995; Grèzes et al., 2004; Happé et al., 1996; Saxe and Kanwisher, 2003). The intuitive notion of “perspective” has, however, a very broad and changeable meaning, which makes it difficult to compare the different studies in this regard. In particular, we like to draw attention to a distinction between different levels of perspective taking skills identified in developmental psychology.

Starting with Piaget and Inhelder’s (1948) Three-Mountains problem, a focal finding was the discovery by Flavell et al. (1981) and Masangky et al. (1974) of the distinction between Level 1 and Level 2 Perspective Taking. Level 1 refers to the ability to distinguish between what people can and cannot see, e.g., that people who look at different sides of a piece of paper see different things: a picture of a cat on the one and a picture of a dog on the other side. Level 2 refers to the understanding that, when people look at the same drawing or scene from different angles, they arrive at different and contradictory descriptions. Fig. 1a gives an example (not the one that Flavell and colleagues had used but material from our study). From our vantage point as readers, we see the pole behind the block, while the little doll in the picture sees the pole to the left of the block. There is clearly a difference of how we and the doll should describe what we see (the speech bubble illustrates that, if the doll gives the description that fits our view, “the block is in front of the pole”, his statement is wrong as a description from his point of view), even though we both look at the same scene and describe the spatial relationship between the two objects.

Level 1 emerges well before Level 2, which, interestingly, is not mastered before about 4 years when other perspective problems are understood. In particular, children become able to understand the false belief problem (Wimmer and Perner, 1983), for example, a person, who fails to witness an unexpected transfer of an object to a new location, will mistakenly believe that the object is still in its original location. False belief problems have become the central original location. False belief problems have become the central...
measure of children’s understanding of the mind (“theory of mind”). They also play a central role in neural imaging of theory of mind in adults.

Flavell’s distinction between two levels of perspective taking illustrates an important ambiguity in the concept of perspective. The word perspective has its origins in the visual arts (Thompson, 1995) of giving the correct visual impression of spatial relationships among objects. Importantly, one talks of people “giving or having a different perspective” only when a particular object or scene is depicted or perceived differently from different points of view. We do not speak of a difference in perspective if different objects or scenes are depicted or perceived. Ambiguity arises because it is, to some degree, at the beholder’s discretion of stipulating whether two pictures depict the same scene (or object) from different points of view or whether they depict different scenes or different parts of a scene.

This ambiguity is very patent in the Level 1 task: one can say that the observers look at the same piece of paper but have different visual impressions (different mental depictions), which justifies calling it a “perspective task”. However, one can with equal justification say that observers look at different sides of the piece of paper (Observer 1 looks at the side with the dog, Observer 2 looks at the side with the cat), hence their differing visual inputs can be explained by a difference in what they are looking at. In contrast, in the Level 2 task, it is difficult to argue that the people are looking at different things. The only difference is the way in which they look at it, which must be responsible for the different interpretations they give: Observer 1 (we as readers) sees the pole behind the block, while Observer 2 (the doll) sees the pole to the left of the block.

An obvious explanation for why children find the Level 2 task more difficult is that this task cannot be understood without an understanding of perspective differences, whereas the Level 1 task can be mastered without this understanding. In other words, some “perspective tasks” involve an analysis of a difference in perspective only optionally (depending on how one conceives of the unity of objects or scenes under consideration), while others (perspective tasks in the strict sense) definitely require an understanding of perspective, in the sense of seeing or depicting, describing one and the same thing differently. Following Perner et al. (2003), we call only the latter "perspective tasks" in the sense of unambiguously requiring some analysis of perspective, which remains optional in the Level 1 tasks.

Clearly, false belief tasks require an understanding of perspective. For instance, when an object is in a new location but the other person believes that the object is still in its original location, then, evidently, the child herself and the other person are not related to different parts of the world (objects, scenes, facts) but conceive of one and the same state of affairs differently: the child conceives of the object as being in location 2, while the other person conceives of it as being in location 1.

If we ask whether there is a brain region associated with registering perspective differences, we might look at the several imaging studies reporting specific areas being activated during
theory of mind tasks, especially those with stories involving false beliefs or ignorance. Meta-analyses of theory of mind imaging studies (Frith and Frith, 2003; Gallagher and Frith, 2003) show consistent involvement of three regions in theory of mind tasks: (1) the anterior cingulate/paracingulate cortex as part of the medial prefrontal cortex, (2) the posterior superior temporal sulcus (pSTS) and (3) the temporal poles. These authors single out the medial prefrontal paracingulate area as particularly responsible for the “decoupling” mechanism (Leslie, 1987) that quarantines representations of imaginary circumstances from straight representations of reality. Such decoupling is, of course, required for representing mental states since the representations of what people believe or want to be the case need to be kept separate from what one knows to be the case. Naturally, this mechanism must be central to representing any perspective that differs from one’s own representation of reality, among them, differences in visual perspective. The importance of this prefrontal region for theory of mind is also underlined by studies of patients with lesions in prefrontal areas mostly including medial areas, who had severe problems attributing first- and second-order false beliefs (Rowe et al., 2001), and problems detecting simple deceptions, especially when medial prefrontal areas are affected bilaterally and on the left (Stuss et al., 2001). Furthermore, the amount of ventromedial frontal atrophy in patients with frontotemporal dementia predicts impairment on theory of mind tests (Gregory et al., 2002).

Recent findings on theory of mind competence from patients with lesions pose, however, some problems for this theory of the anterior paracingulate area. Patient G.T. (Bird et al., 2004) has no demonstrable theory of mind deficit despite extensive damage to the medial frontal lobes bilaterally including the paracingulate area. Apperly et al. (2004) and Samson et al. (2004) report that a group of patients with lesions in the left temporo-parietal junction (TPJ) show specific theory of mind deficits, while patients with medial frontal damage show a mixture of theory of mind and other cognitive and executive impairments. Although these theory of mind impairments cannot be explained as consequences of executive problems caused by prefrontal damage (Rowe et al., 2001), these findings, nevertheless, give some support to the claim by Saxe and Kanwisher (2003) that theory of mind tasks specifically activate the TPJ at the border between the superior temporal and angular gyrus, which is in close vicinity but somewhat dorsal from the pSTS. In summary, there seems to be consensus that the temporal pole is involved in theory of mind processing for other reasons (e.g., social script knowledge, Gallagher and Frith, 2003) but controversy as to whether it is the prefrontal paracingulate area or the posterior STS/temporo-parietal junction area that is centrally involved in theory of mind.

Little noticed, studies of visual perspective taking show no systematic activation of the paracingulate area. To our knowledge, there are only two imaging studies of direct relevance.1 Vogele et al. (2004) looked at how many marks in a room another person (avatar) standing in that room can see in contrast to how many the participants can see looking at the room from outside. One would think that, in particular, judgments about how many marks the avatar can see (3rd person judgment) involve a theory of mind and engage representations of a perspective difference. In this study, the paracingulate part of the medial prefrontal area bilaterally only showed specific activations of 1st person over 3rd person judgments, but no activations for 3rd person judgments either in contrast to 1st person judgments or to the baseline activation. In the vicinity of the posterior STS bilaterally—close to the part of the temporo-parietal junction, where Saxe and Kanwisher reported specific ToM activation—there, too, we find only activations of 1st over 3rd person judgments (we will attempt an explanation of this seemingly contradictory finding in Discussion) or deactivations during 3rd as well as 1st person judgments compared to baseline.

These findings are puzzling for both claims of where theory of mind is computed. In particular, the 3rd person judgments seem to clearly require attribution of the mental state of seeing and computation of the content of this state. Yet, neither the medial prefrontal paracingulate region nor pSTS/TPJ region seems to be involved, despite the impressive consistency of their involvement in other theory of mind tasks involving false belief or ignorance.

One explanation for why neither of these regions is activated by 3rd person judgments is that the perspective problem used in this study is a Level 1 problem, which leaves it open as to whether participants did or did not compute a difference of perspective. The task can be solved without computing perspective differences because the problem can be analyzed in terms of what the avatar is looking at—the two marks in front of him—as opposed to what the participant is looking at and need not be analyzed in terms of avatar and participant looking at the same part of the room from different vantage points leading to different visual impressions of what is in that part of the room. In that case, the theory of mind requirements become minimal because, without representation of visual impressions (perspectives), the seeing can be understood as a spatial relationship between eyes and targets, and this may be the reason why these Level 1 perspective tasks do not activate the “theory of mind” areas in the brain. These regions have been found activated in other theory of mind studies because the problems posed in these studies require a deeper understanding of the mind (see Discussion).

The other imaging study potentially involving visual perspectives is a mental rotation experiment that systematically contrasted viewer rotation with object rotation (Zacks et al., 2003). We focus on viewer rotation because object rotation instructions (as used also in other mental rotation imaging studies) do not create a perspective difference since participants are asked to imagine the array in a different position at a different time. In contrast, viewer rotation instructions ask participant viewers to imagine moving themselves to another vantage point and how the array would look from there, in this particular case, judge whether a particular element in an array would be to the right or left of the viewer’s imagined position. Provided that the participants do compute a view of the static array from a different vantage point, it clearly requires representation of the same scene from a different perspective and ought to engage theory of mind centers dealing with perspective differences.

Nevertheless, viewer rotation (but also object rotation) resulted in bilateral deactivation (against baseline) of the anterior medial prefrontal cortex. Of particular interest are the activation differences between viewer and object rotation found in two areas (although in one case it was in terms of less deactivation under viewer than object rotation), and both of them were within (or close by) the posterior STS area indicated by Frith and Frith’s review of ToM studies.

However, there is no guarantee that participants solved the task by imagining a different view of the array. In fact, this was not even

---

1 There is also a lesion study using visual perspective tasks by Stuss et al. (2001). They found inconsistent impairment in prefrontal patients on visual perspective tasks requiring judgments of whose advice to follow; the person who could or who could not see the hiding of an object. Moreover, unlike the ability to detect deception which was related to damage in medial prefrontal areas, visual perspective taking was not correlated with any particular prefrontal area.
asked for in the instructions. Participants only had to judge whether a target object would be to the left or right of their imagined position. Thus, it is perfectly possible that this task was approached as a pure spatial transformation task without considering any difference in views. The lack of activation in the medial prefrontal paracingulate area, therefore, fails to provide strong evidence against the theory that this area is responsible for theory of mind.

In order to settle such uncertainties about whether alternative visual perspectives have to be computed or not, our study aimed at making representation of contrasting perspectives a necessary feature of the task. In a sentence verification task, participants are shown a visual display of two objects, a squat block and a tall pole and a little doll as an observer figure. In a speech bubble, the observer makes a statement about the spatial relation of the two objects using the perspective relative expressions “in front” and “behind”, e.g., “The block is in front of the pole,” and participants had to judge whether this statement is correct from the viewer’s point of view. Since the viewer is positioned at different angles, his perspective differs from that of the participants’. This condition was contrasted with three comparison conditions resulting in four conditions (shown in Figs. 1a to d) according to a two factorial design: Point of View (Self vs. Other) and Perspective Dependence (a perspective relative spatial relation between objects vs. a perspective independent comparison of object properties).

The prime objective is to see whether the tasks in which a difference in perspective has to be represented activate at least one of the areas activated by false belief tasks, in particular, the anterior paracingulate region and the pSTS/TPJ. There are two possible activation patterns that speak for the representation of a perspective difference. (1) Condition (a: spatial relation for other) requires representation of a perspective difference. Hence, if condition (a) activates a particular region more strongly than the other three conditions, then this region is associated with mandatory computation of a perspective difference. (2) Although not necessary, conditions (b) and (c) are likely to evoke thoughts about different perspectives. Condition (b: spatial relation for self) does so because the spatial descriptions “in front” and “behind” immediately make one aware that these are perspective relative descriptions that raise the danger of a clash of perspectives. Condition (c: property comparison for other) makes perspectival considerations likely because participants are instructed to judge the doll’s statement from the doll’s point of view (even though ignoring the perspective difference would still give the right answer to the question). Only condition (d: property comparison for self) makes any concerns about perspective differences unlikely. Hence, if conditions (a + b + c) activate a particular region more strongly than condition (d), then this region is associated with the computation of likely perspective differences.

In case that one of the ToM regions shows activation pattern (1), i.e., condition (a) activates more strongly than conditions (b + c + d), we added a camera condition (Fig. 1e) in which participants had to judge whether the photo shown at the top of the display could have been taken by the camera from its position in the display. This is to test whether the activation by condition (a) concerns perspective differences in a theory of mind (triggered by the presence of the doll) or whether it is due to perspective differences without involvement of an animate agent and, presumably, not specific to theory of mind. Condition (E: comparison of scenes) served as an additional control for condition (e) because the duplication of scenes (actual scene, photo of scene) in (e) was not part of conditions (a to d).

Materials and methods

Participants

Eighteen volunteers (10 women) were recruited at the University of Salzburg by advertising and were paid 10 for their time. Participants were screened for neurological disorders and contradictions for MRI scanning. The average age was 28.5 years ranging from 21 to 55 years.

Task procedure

Before the scanning took place, the different problem types were explained in detail. Volunteers went through the entire experimental procedure to ensure that, during the presentation in the scanner, no unforeseen problems will be encountered.

Each trial consisted of the presentation of a picture followed by a short fixation cross, which prompted the participants to prepare to answer. There were six different conditions (Fig. 1). In the four sentence verification tasks, participants had to decide whether a written statement placed above a visual scene was true or false of that scene as shown (conditions b and d) or as viewed by the human figure in the display (conditions a and c). In the camera task (condition e), participants had to judge whether the picture shown above the scene could have been made by the camera from its position within the scene. In the comparison of scenes task (condition f), participants had to judge whether the two groups of objects were composed of the same or different kinds of objects. Participants were trained to respond by pressing the left button on a response box if the statement or photo showed the appropriate correspondence with the scene (or if the two scenes in condition f were composed of the same objects) or press the right button otherwise. In every condition, half of the tasks required a left and the other half a right button press.

Design

The sentence verification tasks were organized in a 2 × 2 factorial design with factors Viewer (Self vs. Other) × Perspective Dependence (perspective relative spatial relation vs. perspective independent property comparison). Examples of the 4 resulting conditions are shown in Figs. 1a to d, the camera condition and the scenes-comparison condition in Figs. 1e + f.

Except for condition (f), all visual scenes consisted of a short red block and a tall black pole. Four spatial arrangements of these two objects (from volunteers’ view) were used: block in front, left, right of the pole or behind it. In each of conditions (a to e), each of these four spatial arrangements occurred twice. Depending on condition, these arrangements could be combined with different statements/photos and locations of the human figure/camera. Due to the large variation of possible combinations in each condition, only a set of 8 items of each condition was selected and reused in the second half of the experiment. The following variations were implemented:

(a) In the spatial relation for other condition, the figure could be looking at one of the above described scenes from either the left, right or from behind the scene. The position in the front was never used because then the figure would have had the same view as the participant. Eight combinations of location of figure and spatial relation of the objects were assigned.
randomly to one out of four statements. The statements began either with “the block” or “the pole” and used either the relation “behind” or “in front of”. For each of the two tasks with the same arrangement of objects, one sentence was true and the other false from the perspective of the figure.

(b) In the spatial relation for self condition, no figure was used, leaving 16 combinations of statement and object relation. With the same restrictions as in (a), eight items were selected so that half the sentences were true and the other half false from the perspective of the participants.

(c) In the property comparison for other, the statement used half the time the brightness dimension (lighter–darker) and the other half the size comparison (smaller–taller). The resulting four true and four false sentences (from the figures perspective) were randomly assigned to the scenes used in (a).

(d) In the property comparison for self, the same assignment procedure as in (c) was used. Due to the absence of a figure, it was restricted to the four objects relations so that half the sentences were true and the other half false from the perspective of the participants.

(e) The assignments in the camera condition were the same as in (a) by replacing the human figure with the camera and the statements with the corresponding photos.

(f) In the comparison of scenes condition, the scenes could be composed of an apple, a yellow block, a smaller red block and a black bar. Half of the presented scenes consisted of the same two objects—picked at random—in the left and in the right set. The other half was composed with at least three different kinds of objects. The left-right arrangement of each group of objects was randomized.

In order to be able to compute baseline comparisons for each condition, a baseline activation measure was obtained when participants were looking at a fixation cross for the same length of time as each condition. The same number of such prolonged fixation trials was administered as there were trials of each condition, a baseline activation measure was obtained when participants were looking at a fixation cross for the same length of time as each condition. The same number of such prolonged fixation trials was administered as there were trials of each condition. The same number of such prolonged fixation trials was administered as there were trials of each condition.

**Data acquisition**

MRI data were acquired on the Philips Gyroscan NT 1.5 T scanner (Philips Medical System Inc., Best, the Netherlands) of the Christian Doppler Clinic Salzburg.

Structural MR images were acquired with a MPRAGE sequence (repetition time TR = 17.72 ms, echo time TE = 3.4 ms, FoV = 220 mm, flip angle 12°, 256 × 256 matrix, 130 slices per volume, slice thickness 0.9 mm, voxel size = 0.9 × 0.9 × 0.9 mm³). A supplementary structural image was recorded with the resolution of the functional images. After anatomical imaging, 4 runs of functional scans were obtained. Changes in blood oxygenation level-dependent (BOLD) T2*-weighted MR signals were measured using a gradient echo-planar imaging (EPI) sequence (repetition time TR = 2.3 s, echo time TE = 40 ms, FoV = 220 mm, flip angle 86°, 64 × 64 matrix, 21 slices per volume, slice thickness 6 mm (no gap), voxel size = 3.44 × 3.44 × 6 mm³). For each run, a total of 145 EPI volume images were acquired along the AC–PC plane. The first 6 images were acquired before the beginning of the task to allow transient signals to diminish.

During one functional run, participants encountered each condition two times. The blocks lasted 14 s followed by a variable inter-trial interval and consisted either of two tasks of the same condition or a fixation cross as baseline.

**Image analysis**

Image processing was carried out using SPM2 (Wellcome Department of Imaging Neuroscience, London, UK), implemented in MATLAB 6.5 (Mathworks Inc., Sherborn, MA). Motion correction was done by realigning all functional images to the first scan in the run. Subsequently, all functional images and high-resolution structural scans were co-registered. The low-resolution scan was used for better registration of the functional images to the high-resolution structural scan. These co-registered images were normalized to the Montreal Neurological Institute stereotaxic template brain using nonlinear basis functions. The normalized images of 3 × 3 × 3 mm³ were smoothed by a 9-mm FWHM isotropic Gaussian kernel.

**Results**

**Reaction times**

There were no significant differences in reaction time among the four sentence verification tasks: F(3,45) = 1.238, P > 0.05, (a) mean = 3.58 s, SD = 0.67 s, (b) mean = 3.43 s, SD = 0.61 s, (c) mean = 3.51 s, SD = 0.65 s, (d) mean = 3.44 s, SD = 0.49 s Although we expect these conditions to lead to different brain processes, this lack of differences in reaction time is not surprising since there are no expectations of how long these different processes should take.
Furthermore, reaction times in the camera condition (mean = 3.35 s, SD = 1.23 s) did not differ from the sentence verification tasks. Only the reaction time in the scene comparison control was significantly faster than mean reaction time of the sentence verification conditions (mean = 1.91 s, SD = 0.83 s, t(15) = 13.32, P < 0.001). Already in pilot work was this task perceived as clearly easier.

**Volume of interest**

We first check for the sentence verification tasks whether our perspective conditions (a + b + c) activate part of the two regions that have been consistently found to be activated by ToM tasks involving perspective differences (e.g., false belief tests). For the anterior paracingulate region, the results are very clear. None of our conditions (a + b + c) shows any activation in those regions in comparison to condition (d) at $P_{\text{uncorrected}} \leq 0.001$ at voxel level, except for a single voxel in the comparison (a) with (d) at (−12, 33, 12) that lies close to this region, with $P_{\text{uncorrected}} > 0.60$ at cluster level.

This result confirms the impression from other visual perspective imaging studies (Vogeley et al., 2004; Zacks et al., 2003) that visual perspective tasks do not activate the medial prefrontal areas in the anterior paracingulate, deemed specifically responsible for ToM (Frith and Frith, 2003). The fact that ToM tasks very reliably activate the paracingulate region but visual perspective tasks consistently fail to do so leaves two options: either this area is not responsible for all ToM tasks or visual perspective tasks do not require a theory of mind.

In contrast to the medial prefrontal areas, our perspective conditions do show activations in the posterior regions of the superior temporal sulcus (pSTS) and the temporo-parietal junction (TPJ). We defined a volume of interest to see whether any of these activations are in the vicinity of those in previous studies, in which activations are in the vicinity of those in previous studies, in which clear perspective differences were required, i.e., taking into account another person’s false belief or ignorance. We report any significant cluster of at least 4 voxels in the 2 × 2 factorial design (self vs. other, object comparison vs. spatial relation). These results are shown in Table 1.

The rightmost signal plot in Fig. 2 also shows activations of the camera (e) and the scene comparison (f) conditions in order to see whether these regions are also responsible for computing perspective differences when a camera (instead of another person) is involved. At the location of the interaction effect, the evidence is ambivalent. The camera condition which requires computation of a perspective difference activates this location as much as any of the sentence verification conditions (a, b or c) which require computation of a perspective difference. The problem lies with the scenes comparison control condition, which is only slightly lower than the camera condition ($t(17) = 2.41$, $P = 0.0136$). However, if we compute the contrast between all four conditions (a + b + c + e) that do and the two conditions (d + f) that do not require computation of a perspective difference, that contrast is highly significant at this location: $t(17) = 5.52$, $P < 0.0001$. The activation differences between camera (e) and comparison of scenes (f) in the other two signal plots in Fig. 2 are significant: both $t(17) > 3.37$, $P < 0.002$. Furthermore, the contrasts between conditions that do (a + b + c + e) and those that do not (d + f) require computation of a perspective difference are significant: both $t(17) > 4.05$, $P \leq 0.0004$.

Clearly, this pattern of activations does not speak against the camera task activating the same regions as the sentence verification tasks that compute a perspective difference. However, as evidence for this claim, the data are very weak because—unlike the 2 × 2 design of the sentence verification tasks—the contrast between camera and scenes is too crude to specifically measure computation of a perspective difference.

We also checked where else in our VOI the camera condition (e) contrasts with the comparison of scenes (f). Table 1 shows two such points: A small cluster of specific activation on the right side (Effect 4a) at coordinates closely corresponding to those of the interaction effect (3) at the left side. The signal change values at peak voxel of the four sentence verification tasks show a pattern (with values of 0.9, 0.6, 0.8 and 0.4 for conditions a, b, c and d respectively) reminiscent of the pattern indicative of perspective differences. Contrasting conditions that do (a + b + c + e) with those that do not (d + f) require computation of a perspective difference yields a significant result: $t(17) > 4.68$, $P \leq 0.0001$. This is some evidence that visual perspective differences do not just activate left but also right TPJ similar to false belief tests where activation is often reported bilaterally. Some of the most recent brain imaging studies of false belief reasoning report equally strong (Saxe and Kanwisher, 2003) or even stronger (Saxe and Wexler, 2005) activation in the right than in the left TPJ, but some of the earlier imaging studies also reported uniquely left hemispheric activation (Fletcher et al., 1995; Goel et al., 1995; Happé et al., 1996) or at least stronger activation in the left than in the right hemisphere (e.g., Ruby and Decety, 2003, Grèzes et al., 2004), and also Apperly et al., 2004 report in a lesion study theory of mind impairments predominantly for lesions of the left hemisphere.

Finally, Table 1 shows another significant contrast between camera and comparison of scenes (Effect 4b) at the extreme occipital border of our VOI and clearly the anterior part of a much larger activation of 6602 voxels in basic visual processing areas with its border of our VOI and clearly the anterior part of a much larger activation is often reported bilaterally. Some of the most recent brain imaging studies of false belief reasoning report equally strong (Saxe and Kanwisher, 2003) or even stronger (Saxe and Wexler, 2005) activation in the right than in the left TPJ, but some of the earlier imaging studies also reported uniquely left hemispheric activation (Fletcher et al., 1995; Goel et al., 1995; Happé et al., 1996) or at least stronger activation in the left than in the right hemisphere (e.g., Ruby and Decety, 2003, Grèzes et al., 2004), and also Apperly et al., 2004 report in a lesion study theory of mind impairments predominantly for lesions of the left hemisphere.

**Global analysis**

Table 2 lists all effects for the four sentence verification conditions (a to d) significant at $P < 0.05$ corrected for multiple comparisons at cluster level. Four of these effects in the table pertain to areas in the occipital cortex. The first of them (1) consists

<table>
<thead>
<tr>
<th>Effect/Regions: other &gt; self</th>
<th>Cluster size</th>
<th>MNI coordinates</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parietal inf. + supramarginal + angular</td>
<td>145</td>
<td>−45 −42 39</td>
<td>4.06</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Effect of perspective dependence: relative &gt; independent</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>TPJ (including angular + supramarginal)</td>
<td>6</td>
<td>−48 −48 30</td>
<td>3.49</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Effect of interaction: (b + c) &gt; (a + d)</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>TPJ (extending to supramarginal)</td>
<td>14</td>
<td>−51 −39 33</td>
<td>3.65</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Effect of camera: camera &gt; scenes (cluster size restricted to voxels within VOI)</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Supramarginal</td>
<td>7</td>
<td>−51 −45 35</td>
<td>3.66</td>
</tr>
<tr>
<td>Occipital mid. + temporal mid. + angular</td>
<td>286</td>
<td>−36 −75 21</td>
<td>5.41</td>
</tr>
</tbody>
</table>
of a large cluster in the occipital cortex centered in the right middle occipital gyrus but spreading also into its left parts, and bilaterally into the fusiform, middle temporal, superior and inferior occipital gyrus and the cerebellum. The activations are due to the other–self difference. They most likely reflect the greater amount of visual elements in the other (presence of the doll) than in the self conditions as also reported by Voogeley et al. (2004).

The size of this first cluster highlights a weakness in our stimulus material. The manipulation of the other–self difference with the presence versus absence of a powerful visual stimulus (the doll) makes this contrast difficult to interpret as the effect of considering another person’s perspective. One should therefore keep in mind that it can only be interpreted in this way if it is modulated by an appropriate interaction with the other factor of a perspective relative versus perspective independent statements.

Two further effects (4 + 5) are due to object comparison (c + d) vs. spatial relations (a + b) with significant bilateral activations in middle occipital gyrus and calcerine, spreading into inferior occipital and fusiform gyrus and the cerebellum. This is most likely due to the different need for visual analysis of object properties than spatial relations.

Effect 6 shows that the main effect just discussed is bilaterally modulated by a significant interaction. The activation patterns of the four conditions tested against baseline indicate that activation is higher in condition (c) than the other three conditions ($F(3,85) = 1.365, P < 0.05$; (a) mean = 1.84, SD = 1.16; (b) mean = 1.66, SD = 0.93; (c) mean = 2.71, SD = 1.32; (d) mean = 1.52, SD = 1.15). The interpretation of this effect is not obvious. If it were just a summation of visual processing effort stemming from the richer visual input in the other conditions (a + c) and the need for representing object properties (c + d), then we would expect a graded activation for c, a + d and b. One possible explanation is that participants, instructed to judge the truth of the uttered sentence from the speaker’s point of view, make an effort to do so and incorporate brightness (or size) information when they imagine the two objects as seen from the speaker’s view. Hence, only in condition (c) the visual analysis of object properties is carried out twice: once for primary processing the input and a second time for imagining the objects from the doll’s point of view. Importantly, this result then confirms our assumption that participants do compute a perspective for the doll in condition (c)–even though they could ignore it for answering the question correctly—which in turn strengthens our interpretation of the interaction pattern in the pSTS/TPJ region of interest that condition (a + b + c) induces representation of a difference in perspectives.

The difference between other and self also activates the precentral and frontal gyri. In the left hemisphere (Effect 2), the activation is in the superior frontal and the precentral gyrus. This location is also reported by Ruby and Decety (2001) when imagining either the experimenter (3rd person perspective) or imagining oneself (1st person perspective) acting on an object as opposed to a control condition of passively watching a set of objects. Both 1st and 3rd perspective instructions have something in common with our other conditions (a + c). They induce participants to imagine a person looking at the scene, while the self conditions (b + d) do not ask to imagine anyone looking at the scene (like Ruby and Decety’s control conditions). This interpretation is also underlined by activation in this area in the study by Voogeley et al. (2004) contrasting 3rd and 1st person perspective against baseline, where participants were asked how many dots the avatar can see or how many dots they themselves can see.

In the right hemisphere (Effect 3), activation could be seen in the inferior frontal (operculum and triangularis) and in the precentral gyrus. Voogeley et al. (2004) report similar activation in this area for contrasting 3rd vs. 1st person as well as for contrasting 3rd and 1st person perspective taking against baseline. Samson et al. (2005) interpret activation in this region on the basis of a lesion study as stemming from the need to inhibit the self perspective.

The remaining two entries in Table 2 (Effects 7 + 8) show significant interaction effects bilaterally along the Sylvian fissure (transition between superior temporal gyrus and Rolandic operculum). Signal change plots for Effect 7 in the right hemisphere for
### Table 2

Significant global effects for sentence verification conditions

<table>
<thead>
<tr>
<th>Effect/Regions</th>
<th>Cluster size</th>
<th>MNI coordinates</th>
<th>Z</th>
<th>( P )</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Effect of viewer: other &gt; self</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1) Bilateral: occipital + fusiform + cerebellum + temporal mid.</td>
<td>4372</td>
<td>33 −84 18</td>
<td>6.13</td>
<td>0.000</td>
</tr>
<tr>
<td>(2) L: superior frontal gyrus + precentral</td>
<td>82</td>
<td>−24 −18 60</td>
<td>4.09</td>
<td>0.007</td>
</tr>
<tr>
<td>(3) R: precentral + inferior frontal gyrus</td>
<td>76</td>
<td>48 6 27</td>
<td>3.69</td>
<td>0.01</td>
</tr>
<tr>
<td><strong>Effect of perspective dependence: relative &gt; independent</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(4) R: occipital mid. and inf. + lingual + fusiform</td>
<td>180</td>
<td>18 −99 3</td>
<td>4.42</td>
<td>0.000</td>
</tr>
<tr>
<td>(5) L: calcerine + lingual + occipital inf. + cerebellum + fusiform</td>
<td>414</td>
<td>−21 −90 3</td>
<td>5.31</td>
<td>0.000</td>
</tr>
<tr>
<td><strong>Effect of interaction: ((b + c) &gt; (a + d))</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(6) bilateral calcerine + 1 cuneus + r lingual</td>
<td>423</td>
<td>12 −72 15</td>
<td>4.72</td>
<td>0.000</td>
</tr>
<tr>
<td>(7) R: temporal sup. + Rolandioc operculum</td>
<td>103</td>
<td>45 −24 9</td>
<td>4.23</td>
<td>0.002</td>
</tr>
<tr>
<td>(8) L: Rolandic operculum + temporal sup.</td>
<td>66</td>
<td>−51 −18 15</td>
<td>3.87</td>
<td>0.03</td>
</tr>
</tbody>
</table>

The four sentence verification tasks indicate a strong deactivation in condition, (a) while the other three conditions show no systematic deviation from baseline \( F(3,68) = 1.365, P < 0.05; (a) \text{mean} = −0.81, SD = 0.7; (b) \text{mean} = −0.25, SD = 0.46; (c) \text{mean} = −0.21, SD = 0.48; (d) \text{mean} = −0.37, SD = 0.65). Considering that condition (a) requires most clearly realization that the other person’s perspective differs from one’s own, this deactivation fits well with previously reported higher activation for 1st than for 3rd person perspective taking in this region. Ruby and Decety (2003) reported higher activation for making an argument from own knowledge than for making the argument based on others’ erroneous assumptions. Vogeley et al. (2004) reported specific activation for 1st over 3rd person seeing judgments in a region close by in terms of MNI coordinates but whose peak activation was given as located in the insula.

Our results add to these previous studies in that we also find a very similar area activated in the left hemisphere (Effect 8), for which we could not find corresponding results in the relevant perspective taking literature.

### Discussion

The puzzle we started with was why visual perspective tasks, in which participants have to figure out what another person sees, do not seem to activate reliably the cerebral areas deemed necessary for theory of mind. In a review of theory of mind experiments, Frith and Frith (1999, 2003) and Gallagher and Frith (2003) singled out the anterior medial prefrontal paracingulate area as centrally involved in theory of mind tasks. This area failed to be activated when judging what another person can see (Vogeley et al., 2004) and what oneself would see if one changed position (viewer rotation condition in a mental rotation task by Zacks et al., 2003). One possible explanation for this activation failure was that the tasks in these studies did not require participants to compute perspective differences, one of the hallmarks of false belief tasks, frequently used to assess theory of mind. This leaves it possible to give correct answers to questions about what another person sees without having to compute anyone’s perspective. It can be done on the basis of computing spatial relations between the observer and the observed scene (Level 1 perspective tasks). To test this possibility, our perspective tasks required representation of a perspective difference (Level 2 task). Nevertheless, the relevant medial prefrontal area still failed to be differentially activated.

We did find specific activation in the pSTS/TPJ area, which has been claimed more recently to be necessary for theory of mind tasks, in particular false belief tasks (Saxe and Kanwisher, 2003; Apperly et al., 2004) and has been reported by Frith and Frith (1999, 2003) to be consistently activated by theory of mind tasks. Furthermore, the viewer rotation condition used by Zacks et al. (2003) as compared to array rotation showed activation in this region. An interesting question remains, why judgment of what another person sees in Vogeley et al.’s (2004) study did not show activation in this area. As indicated above, this may be because their 3rd person condition did not require and may, therefore, have been solved without representation of a perspective difference. Interestingly, there was activation in this area in their 1st person condition, judging how many target objects oneself can see in a room. Since the participants did not have a complete view of the room, that condition may have made them wonder whether they could see all target objects and how many there might really be. In that case, they were computing a perspective difference between an unknown actual number of targets and the number they could see.

In an attempt to integrate recent brain imaging and lesion evidence on the neuropsychology of theory of mind, we venture the following suggestion exemplified by false belief tasks. To understand false beliefs, at least two elements are required: one has to represent belief as a perspective on the world, and one has to understand the role of the belief in generating action in conjunction with a person’s goals or desires (so-called belief–desire reasoning).

Our suggestion is that the pSTS/TPJ area is responsible for representing different perspectives and for making behavioral predictions (cold facts about the mind). In contrast, the medial prefrontal area is responsible for making behavioral predictions and anticipating the wider consequences of this behavior, in particular, its emotional consequences (hot facts about the mind). This can explain the following findings:

1. Lesions in the pSTS/TPJ area lead to specific inability to carry out false belief tasks (Apperly et al., 2004) because patients cannot represent the difference in perspective. In contrast, lesions in the medial prefrontal area (Patient G.T.\(^2\)) may leave this ability in

\(^2\) The authors did describe G.T. with “a slight blunting of empathy for the characters in the story”, which fits an impairment in the emotional consequences of the beliefs and desires, for which the missing medial prefrontal area is responsible.
contrasted with normal social behavior (Berthoz et al., 2002) and 2000), when intentional and unintentional social violations are or below that region when shown moving shapes (e.g., Castelli et al., 2004, showing intentional actions and interactions between people) moving shapes that lead to attributions of intentions; Walter et al., 2003).

States report activation of medial prefrontal regions but often not of perspectival mental states like belief, but non-perspectival goal parts of our region of interest (e.g., Schultz et al., 2003, showing moving shapes that lead to attributions of intentions; Walter et al., 2004, showing intentional actions and interactions between people) or below that region when shown moving shapes (e.g., Castelli et al., 2000), when intentional and unintentional social violations are contrasted with normal social behavior (Berthoz et al., 2002) and when pretend actions, e.g., acting as if one was putting a book on the shelves, are contrasted with real actions, i.e., putting a book on the shelves (German et al., 2004).3

A large variety of tasks that require attribution of emotions to self or other activate the medial prefrontal area (see Ochsner et al., 2004, for review).

When contrasting strategic games played against a human with being played against a computer pSTS/TPJ is not specifically activated, whereas the medial prefrontal areas are (Gallagher et al., 2002; McCabe et al., 20014). The pSTS/TPJ area is not specifically activated for humans because either no perspective contrast is required or because it is required for playing against humans as well as the computer. In contrast, playing against a human has potential emotional consequences (the partner may resent one’s competitive strategy), which playing against a computer does not have. This difference in the anticipated behavior’s emotional consequences explains the specific activation of the medial prefrontal area.

To answer the question in our title: visual perspective tasks do need a theory of mind, but only a specific part of it. It is an important part, namely, the realization that minds can take different perspectives on the world because they represent it differently. This ability is required for false belief tests as well as visual perspective tasks, and we locate it in the dorsal part of the pSTS/TPJ region. Under the label of “theory of mind,” many other abilities of more practical importance are subsumed, in particular, the ability to predict how organisms with certain goals will act under certain circumstances. In most everyday cases, such predictions can be made without concern that goals and circumstances are represented with a certain perspective by the organism. We locate this ability in the ventral part of the pSTS/TPJ (and regions ventrally adjacent to it) as well as in the medial prefrontal cortex (particularly important for computing emotional consequences). These two levels of a “theory of mind” have correspondences in developmental psychology. Knowledge of how people act without a perspectival understanding is seen as developing into a perspectival understanding of the mind evident from contrastive terms like “Desire vs. Belief + Psychology” (Wellman, 1990), “situation theory of behavior vs. representational theory of mind” (Perner, 1991) or “teleological stance vs. mentalistic action theory” (Csibra and Gergely, 1998).

**Acknowledgments**

The authors are grateful to the members of the Department of Radiology for assistance and thank Chris Frith for his encouragement at the early stages of this research.

**References**


---

3 This last finding that observing pretend action does not activate the same pSTS/TPJ region as belief reasoning is of particular interest because pretence and false belief can be used to explain why people act in a way that does not make sense given the state of the world. However, they differ in that someone’s false belief must be understood as representing the given state of the world by giving a deviant perspective (a mistaken view) of that state, whereas the mental representation guiding a pretend action can be understood as a representation of a possible state of the world and need not be understood as representing the real state of the world and giving a deviant perspective of it (Perner, 1991; Perner et al., 2002).

4 Their Figure 4 does show a tiny spec of specific activation in pSTS/TPJ—but nothing compared to the large and substantial activation in the medial prefrontal area.