Heterogeneity of social cognition between visual perspective-taking and theory of mind in the temporo-parietal junction

Kenji Ogawa & Yuiko Matsuyama

Department of Psychology, Graduate School of Humanities and Human Sciences,

Hokkaido University

Correspondence should be addressed to Dr. Kenji Ogawa, Department of Psychology, Graduate School of Humanities and Human Sciences, Hokkaido University, Kita 10, Nishi 7, Kita-ku, Sapporo, 060-0810 Japan; Tel/Fax: +81-011-706-4093; E-mail: ogawa@let.hokudai.ac.jp

Abbreviated title: Heterogeneity of social cognition in the TPJ

Keywords

fMRI; visual perspective-taking (VPT), theory of mind (ToM), temporo-parietal junction (TPJ)

Acknowledgments

This work was supported by JSPS KAKENHI Grant Numbers 17H01016 & 19H00634 to K.O.

The authors would like to thank Enago (www.enago.jp) for the English language review.

Conflict of interest

The authors declare no competing financial interests.

Code accessibility

The computer code is available from the corresponding author upon reasonable request.

Abstract

Visual perspective taking (VPT), particularly level 2 VPT (VPT2), which allows an individual to understand that the same object can be seen differently by others, is related to the theory of mind (ToM), because both functions require a decoupled representation from oneself. Although previous neuroimaging studies have shown that VPT2 and ToM activate the temporo-parietal junction (TPJ), it is unclear whether common neural substrates are involved in both functions. To clarify this point, we directly compared the TPJ activation patterns of individual participants performing VPT2 and ToM tasks using functional magnetic resonance imaging and within-subjects design. A whole-brain analysis revealed that VPT2 and ToM activated overlapping areas in the posterior part of the TPJ. In addition, we found that both the peak coordinates and activated regions for ToM were located significantly more anteriorly and dorsally within the bilateral TPJ than those measured during the VPT2 task. We further confirmed that these activated areas were spatially distinct from the nearby extrastriate body area (EBA), visual motion area (MT+), and the posterior superior temporal sulcus (pSTS) using independent localizer scans. Our findings revealed that VPT2 and ToM have gradient representations, indicating the functional heterogeneity of social cognition within the TPJ.

Introduction

Visual perspective taking (VPT) is the ability to understand how the world looks to others and has two different levels. Level 1 VPT (VPT1), also called perspective tracking [1], is the ability to determine if another person can see an object, whereas level 2 VPT (VPT2) is the ability to understand that others have different views of an object than one's own [2]. Previous studies have indicated that VPT1 and VPT2 are distinct cognitive processes. Developmental studies have shown that VPT1 is acquired between 18 and 24 months of age, whereas VPT2 appears in 4- or 5-year-old children [3–5]. Behavioral studies also support that VPT1 and VPT2 have different properties. In VPT2, the response times increase with the increase in the angular distance between the participant and the agent, indicating that the participant mentally transformed their own position to the other's perspective [6]. Studies using mental simulation of body movements also proposed that VPT2 is an embodied process [7,8].

VPT2 is related to the theory of mind (ToM), because both functions require a decoupled representation from one's own perspective or belief. ToM or mentalizing refers to the capacity to understand and manipulate other people's behavior based on their mental states [9]. Moreover, people with autism spectrum disorder (ASD) present impaired ToM and VPT2, whereas the VPT1 ability remains intact [3,10]. Hamilton et al. (2009) compared the performance of children with ASD with those of typically developed children in terms of the VPT2 and in mental rotation (MR) task, both of which require comparable mental spatial processing. The difference between VPT2 and MR is that VPT2 requires the participants to take the perspective of other people that is distinct from their own view, while MR requires the imagined rotation of an object without any change in their perspective. They showed that VPT2, but not MR, performances were significantly impaired in the children with ASD compared with those of typically developed children. They also showed that VPT2 performances were

significantly associated with the ToM ability [10]. Their findings indicate that VPT2 and ToM have common cognitive substrates.

The temporo-parietal junction (TPJ) has been involved in both VPT and ToM. Lesions in the TPJ induce impairments of the ToM, particularly of the ability to understand others' false beliefs [11,12] and of VPT, especially the ability to take another person's perspective [13]. In addition, transcranial direct current stimulation of the TPJ improves VPT performances in normal subjects [14,15]. Many neuroimaging studies have consistently shown TPJ activations during the processes of VPT [16–19] and ToM [20–24]. Some meta-analyses have also shown that both VPT and ToM cause TPJ activation [1,25]. However, another study found no overlap activation between VTP and the different types of ToM tasks [26]. Although VPT and ToM activate regions within the TPJ, they might not share the same neural substrates. The activations within the same region of interest do not necessarily indicate common neural substrates because a single brain region (or a single voxel) usually contains multiple subpopulations of functionally distinct neurons that are beyond the spatial resolution of a typical functional magnetic resonance imaging (fMRI) [27]. Since the TPJ is heterogeneous and composed of various subregions [28–30], it is important to investigate the common neural substrates between VPT and ToM by testing the regions activated within individual participants to remove any cross-individual variability.

The present study aimed to investigate whether VPT2 and ToM share common neural substrates in the TPJ using within-subjects fMRI design. Participants performed VPT2 and ToM tasks. The regions activated by VPT2 were compared with those activated by an MR task as the control task. The activities triggered by ToM were identified with a standard ToM localizer. The activations were then compared within participants to identify potential shared neural substrates for VPT2 and ToM in the TPJ. Hereafter, we used the term VPT to indicate VPT2.

Materials and Methods

Participants

Participants included 31 healthy volunteers (21 males and 10 females; 20–27 years of age). The number of participants was determined to detect reliable activations based on previous fMRI experiments on VPT [16,17] or ToM [23,31]. All participants, with the exception of one, were right-handed, as assessed by a modified version of the Edinburgh Handedness Inventory [32] for Japanese participants [33]. Written informed consents were obtained from all participants in accordance with the Declaration of Helsinki. The experimental protocol was approved by the local ethics committee of Center for Experimental Research in Social Sciences at Hokkaido University.

Task procedures

Each participant underwent a VPT/MR and a ToM task. For the VPT/MR task, we employed a computerized version of the paradigm by Hamilton et al. [10] for VPT and MR tasks for children. Both VPT and MR require comparable mental spatial processing; hence, subtracting the activations of MR from those of VPT enables us to reveal activities related to perspective-taking of others. Although their original study used a real meaningful object (small toy), we used pseudo-randomly combined three-dimensional cubes typically used for MR [34] to increase the task difficulty for adults. First, three-dimensional cubes (Sheppard figures) and a pot, both placed on a tray, were displayed on the left and right sides of the screen, respectively. The front side of the tray was always colored red on the left figure, while one of four sides of the tray was randomly colored red on the right figure. Both figures (cubes and a pot) were presented simultaneously within one display for 3 s. In addition, under the VPT condition, an image of a person (avatar) standing either on the front, back, left, or right sides of the pot was displayed. The participants were asked to imagine the view of the object from the perspective of

the avatar. Under the MR condition, the participants were asked to imagine the tray rotating in the direction indicated by the red side of the tray as well as the rotation of the object. Under VPT and MR conditions, four different views of the same object were presented, and the participants had 4 s to select the correct response (Fig. 1). Between the trials, a fixation cross was presented for 8 s. The VPT and MR trials were pseudo-randomly presented in the same run, with 20 trials per condition. One run lasted for approximately 10 min.

---------- FIGURE 1 ABOUT HERE ----------

For the ToM task, we used a modified version of standard ToM localizer [20] for Japanese participants [35]. The participants were firstly presented with stories describing false beliefs in the ToM task or those about false pictures or maps in the control task. Subsequently, they were asked a true or false question for 4 s regarding whether the situation was real or false. The ToM condition requires the outdated "false" beliefs of other people to understand their actions, while the control condition requires the representation of false or outdated content without a person. As an example of a ToM condition, the first story was "It was the morning of the high school reunion. Yuki put her high heels under her clothes and went shopping. That afternoon, her sister borrowed them and later put them back under Yuki's bed." The follow-up question was "When Yuki prepares to go to her high school reunion (at night), she thinks that her high heels are under her clothes." and this requires a true response. As an example of a control condition, the story was "The family's old videotapes showed them celebrating their daughter's first birthday at their home in Nagoya. The family later sold the house and moved to Osaka." The question was "The video shows the family in Nagoya." that requires a true response. For both ToM and the control, the stories were presented with written sentences on the display, so basic sensory inputs were comparable between conditions. Both types of stories were pseudo-randomly ordered within a run, with seven stories per condition. Each story was

displayed as written sentences for 10 s. Between the trials, a fixation cross was displayed for 7 s. One run lasted for approximately 5 min.

VPT/MR and ToM tasks were conducted in separate runs. Each participant underwent a total of three runs, two VPT/MR runs and one ToM run. While VPT/MR tasks were conducted for two runs to gather a sufficient number of trials, we conducted only one run for ToM/control tasks following a standard localizer protocol, which could activate the bilateral TPJ [35]. The order of the VPT and ToM runs was counterbalanced across participants. The participants' response (pressing a button) and reaction time were recorded using an MRI-compatible response pad (Current Design, Philadelphia, USA). The stimuli were designed using SketchUp Make 2017 (Trimble Inc., California, USA), and the programs were implemented with PsychoPy 3 (https://www.psychopy.org/).

Magnetic resonance imaging (MRI) acquisition

All scans were performed on a Siemens (Erlangen, Germany) 3-Tesla Prisma scanner with a 64-channel head coil at Hokkaido University. T2*-weighted echo planar imaging (EPI) was used to acquire a total of 304 scans for the VPT runs and 151 scans for the ToM runs, with a gradient echo EPI sequence. The first three scans within each run were used for T1 equilibration and were discarded. The scanning parameters were repetition time (TR): 2,000 ms, echo time (TE): 30 ms, flip angle (FA): 90°, field of view (FOV): 192 × 192 mm, matrix: 94 × 94; 36 axial slices, and slice thickness: 3.0 mm with a 0.75 mm gap. T1-weighted anatomical imaging with a magnetization-prepared rapid gradient-echo sequence was performed using the following parameters: TR: 2,300 ms, TE: 2.32 ms, FA: 8°, FOV: 256 × 256 mm, matrix: 256 × 256, 192 axial slices, and slice thickness: 1 mm without a gap.

Processing of fMRI data

Image preprocessing was performed using the SPM12 software (Wellcome Department of Cognitive Neurology, http://www.fil.ion.ucl.ac.uk/spm). All functional images were initially realigned to adjust for motion-related artifacts. Volume-based realignment was performed by co-registering images using rigid-body transformation to minimize the squared differences between volumes. The realigned images were then spatially normalized using the Montreal Neurological Institute (MNI) template based on the affine and non-linear registration of coregistered T1-weighted anatomical images (normalization procedure of SPM). They were resampled into 3-mm-cube voxels using sinc interpolation. Images underwent spatial smoothing using a Gaussian kernel of 6 × 6 × 6-mm full width at half-maximum.

Using the general linear model, the task blocks of each session were modeled as box-car regressors that were convolved with a canonical hemodynamic response function. For both VPT and ToM runs, the box-car covered the presentation period of the first stimuli (4 s for VPT and 10 s for ToM). The six realignment parameters were also included in the design matrix as covariates. Low-frequency noise was removed using a high-pass filter with a cut-off period of 128 s, and serial correlations among scans were estimated with an autoregressive model implemented in SPM12.

fMRI analysis

We used conventional mass-univariate analysis of individual voxels to identify the activated areas. For VPT runs, we analyzed areas that were significantly activated during the VPT task compared with those activated during MR trials (VPT > MR). For ToM runs, the regions that were significantly activated during the ToM task were compared with those activated in control trials (ToM > control). Contrast images of each participant, generated using a fixed-effects model, were analyzed using a random effects model of one-sample t-test. The activation was reported with a threshold of $p < .05$ corrected for multiple comparisons of family-wise error (FWE) at the cluster-level and with $p < .001$ uncorrected at the voxel level. In addition, to

compare their activation patterns, individual voxel coordinates of peak activation were selected within a spherical region of interest (ROI) with a radius of 20 mm centered on the reported coordinates of the bilateral TPJ (right TPJ: –49, –61, 27; left TPJ: 54, –55, 20) based on a recent meta-analysis of neuroimaging studies for social cognition [36]. The same ROI mask was applied for VPT and ToM analysis to ensure unbiased selection of the coordinates. To compare individual peaks within the TPJ, peak voxels were selected within the TPJ ROI mask at $p < .05$, uncorrected for the voxel level.

Functional localizer scans

We additionally conducted standard functional localizer scans to identify the following three regions in each participant: the extrastriate body area (EBA) [37,38], visual motion area (MT+) [39,40], and the posterior superior temporal sulcus (pSTS) [41,42]. These three localizer runs were independently scanned for the 12 subjects (9 males and 3 females; 20–24 years of age) who participated in the main experiment. Each run lasted for approximately 7 min.

In the EBA localizer, one block consisted of 20 images of headless human body parts in different postures, which were alternated with 20 images of chairs as controls. Each image was presented for 300 ms, followed by a black screen for 450 ms. A fixation cross was intercalated at the end of each block for 12 s as baseline. The same images were presented twice in succession, twice during each block. The participants were asked to press a button with the right index finger when they detected the immediate repetitions. In the MT+ localizer, the viewing block consisted of presenting moving or static dots for 15 s alternately with a rest period of fixation display for 12 s. A total of 16 blocks, which consisted of 8 moving blocks and 8 static blocks, were executed. In the pSTS localizer, a point-light biological motion was displayed for 15 s. In the control condition, scrambled displays with the same motion vectors as those of the biological motion but with randomized initial starting positions were presented for 15 s. A total

of 16 blocks were performed, which consisted of 8 biological motion blocks and 8 control blocks. The participants passively viewed the stimuli during the EBA and pSTS localizer scans. The threshold of the activation peak was set at $p < .05$ uncorrected for the multiple comparisons and selected within a spherical mask with a 20-mm radius centered on the previously published coordinates as follows: EBA coordinates were based on the mean coordinates of nine neuroimaging paper [43], whereas MT+ and pSTS were selected from the meta-analysis of social cognition [36].

Results

Behavioral analysis

We analyzed the task accuracy and the reaction time (RT) (Fig. 2). The mean (SD) accuracies in the VPT and MR tasks were 87.3% (14.5%) and 70.6% (20.0%), respectively. Results of the paired t-test showed that the accuracy in the VPT task was significantly higher than that in the MR task ($t(30) = 5.07$, $p < .001$, Cohen's $d = 0.93$). The mean (SD) RTs in the VPT and MR tasks were 2.63 (0.59) and 2.71 (0.56) s, respectively. The VPT RTs were significantly shorter than those observed in the MR task ($t(30) = 2.20$, $p < .05$, Cohen's $d = 0.40$).

The mean (SD) accuracy in the ToM localizer task and control condition were 80.6% (19.0%) and 85.7% (16.5%), respectively, without significant difference between both tasks ($t(30) = 1.61$, $p = 0.12$, Cohen's $d = 0.29$). The mean (SD) RTs in the ToM localizer task and control condition were 4.03 (1.07) and 3.19 (1.07) seconds, respectively. The RTs in the ToM localizer task were significantly slower than those in the control condition ($t(30) = 5.08$, $p < .001$, Cohen's $d = 0.93$).

---------- FIGURE 2 ABOUT HERE ----------

Analysis of fMRI data

A whole-brain analysis was conducted to identify the activations specific to the VPT and ToM conditions. We found notably higher activations in the bilateral TPJ during VPT tasks than in those observed in MR trials (VPT > MR). The regions preferentially activated during ToM compared with those activated in control condition were mostly in the bilateral TPJ extending into the anterior temporal cortex (ToM > control). Activations of the medial prefrontal cortex (mPFC), the precuneus, the early visual cortex, and the cerebellum were also detected (Fig. 3A, B, and Table 1). We then overlapped the voxels activated during the VPT and ToM tasks and

found a relatively small overlap (blue voxels in Fig. 3C and D, number of voxels = 69) compared with the original cluster size activated by VPT (Fig. 3A in green, number of voxels = 130) and ToM (Fig. 3B in red, number of voxels = 3,124).

---------- FIGURE 3 ABOUT HERE ----------

To compare individual peaks within the TPJ, peak voxels were selected within the TPJ ROI mask. We selected the voxel that showed the highest t-value within the ROI to extract the peak coordinates. No significant activation ($p < .05$, uncorrected for the voxel level) was found within the TPJ in the left and right hemispheres of 3 and 2 participants, respectively, during the VPT or ToM tasks. Therefore, these participants were excluded from the coordinate comparison. The peak coordinates were plotted using BrainNet Viewer [44]. The peak voxels obtained during the ToM task were found in locations significantly more anterior (y-axis left hemisphere: $t(27) = 3.06$, $p = .005$, Cohen's $d = 0.59$; y-axis right hemisphere: $t(28) = 2.70$, $p = .012$, Cohen's $d = 0.51$) and dorsal (z-axis left hemisphere: $t(27) = 3.52$, $p = .002$, Cohen's $d = 0.68$; z-axis right hemisphere: $t(28) = 2.49$, $p = .019$, Cohen's $d = 0.47$) compared with the peak voxel positions in the bilateral TPJ for VPT tasks. There was no significant difference in the lateromedial direction (x-axis left hemisphere: $t(27) = 0.85$, $p = .40$, Cohen's $d = 0.16$ and x-axis right hemisphere: $t(28) = 0.53$, $p = .60$, Cohen's $d = 0.10$) (Fig. 4).

---------- FIGURE 4 ABOUT HERE ----------

To test the differences in the voxel intensities between VPT and ToM, we also directly compared the activation images (beta maps) between two conditions within a single statistical model (design matrix) using SPM12. This analysis includes the interactions between VPT and ToM with respect to corresponding control conditions: (VPT > MR) > (ToM > ToM control) and vice versa. The activities were inclusively masked with the same TPJ-ROI used for the analysis of individual peak coordinates: a sphere with a radius of 20 mm centered on the

reported coordinates [36]. This analysis showed that the activated regions by ToM were located more anteriorly and dorsally within the bilateral TPJ than those measured during the VPT ($p < .05$, uncorrected for the voxel level; cluster threshold = 10 voxels) (Fig. 3E). This result is congruent with our previous results based on individual peak coordinates and corroborates the heterogeneity of VPT and ToM within the bilateral TPJ.

We finally used independent functional localizer scans to determine the peak localization in each participant for the EBA, MT+, and pSTS. All participants showed reliable EBA and MT+ activations in both the hemispheres, whereas we observed no activation in the pSTS from the left and right hemispheres for 5 and 6 participants, respectively, and therefore did not include them in the analysis. We compared the individual peaks and found that these functional clusters were significantly spatially distinct from those obtained during VPT and ToM tasks in the y- and z-axes for both hemispheres (Fig. 4) ($ts > 2.32$, $ps < 0.05$, Cohen's $ds > 0.73$).

Discussion

This study investigated whether the same neural substrates in the TPJ are involved in VPT and ToM. We found focal activities in the bilateral TPJ induced by VPT compared to MR-related activation. This TPJ activation was partially overlapped with the ToM activity. Previous studies showed that the reorientation of the attention activates the TPJ [45]. There have been controversies about whether the same TPJ areas are involved in the reallocation of attention and the ToM [29,31,46–48]. A meta-analysis revealed a substantial overlap between attentional reorientation and ToM in the right TPJ (rTPJ) [29], whereas other studies showed that the posterior part of the rTPJ is involved in social cognition [49,50]. The TPJ activations observed in the present study are unlikely owing to the attentional processes, because VPT and MR require bottom-up attention to visual cues and comparable mental spatial processing [10]. The behavioral analyses showed faster and more accurate responses in VPT than in MR, which might indicate that the attentional demand or the global task difficulty was lower in VPT than in MR. In addition, the TPJ activation peak was closer to the coordinates obtained for ToM than those in attentional reorientation [29]. For example, the coordinates [54 –64 14] in the right hemisphere (Table 1) are closer to the mean coordinates found in ToM [56, –54, 19] than those obtained in previous studies for attentional reorientation [55, –55, 26] [46,50]. The meta-analysis has revealed both dorsal and ventral clusters around the left TPJ for the common activation between VPT and ToM (false belief) (Arora et al., 2017). In contrast, the current study only revealed the ventral cluster for VPT compared with the MR condition. This discrepancy could be due to our contrast of VPT vs. MR, considering that VPT and MR require viewer rotation and object rotation, respectively, which are subserved by different neural substrates in the posterior parietal cortex [51]. It should also be noted that the presence of a person was confounded with VPT vs. MR comparison, which could induce the activation of the region involved in visual body perception, including EBA [37]. To overcome this problem, we

additionally conducted localizer scans to individually identify the EBA and found that the EBA are located distinctly with the current activities.

We found not only large areas activated by ToM, including the bilateral TPJ, but also the anterior part of the temporal cortex, the mPFC, the precuneus, and the lingual gyrus. Although previous studies indicated a dominance of the right hemisphere in ToM [50,52,53], a bilateral involvement was also reported [23,54], which is consistent with our result. The large area of activity in both hemispheres also agrees with previous neuroimaging studies on ToM [35,55,56], constituting a core network for ToM performance [9,22,57,58].

By comparing the whole-brain activations during VPT and ToM, we found partially overlapped voxels within the TPJ. This overlap was located within the angular gyrus and corresponded with the posterior part of the TPJ (TPJp), rather than the anterior TPJ (TPJa), which is consistent with the roles of the TPJp in social cognition, whereas the TPJa is involved in attentional reorientation [28,29,50,59,60]. We further compared the peak activations of both functions in individual participants and found that the peak voxels induced by ToM were located significantly more anteriorly and dorsally within the bilateral TPJ compared with those observed during the VPT task. Both VPT and ToM share cognitive processes that require the mental representation of other agents, which is decoupled with the self-perspective for VPT or one's own belief for ToM. Additionally, ToM necessitates a more complex representation of one's beliefs or thoughts for a longer time during narrative comprehension, whereas VPT requires a more visual and instantaneous representation of another's perspectives. This functional discrepancy in representing one's minds might cause distinct peaks of activities during ToM and VPT.

The VPT and ToM activity peaks were also found more anteriorly and dorsally than activities in the EBA and pSTS, both of which are involved in social perception of others [37], and the pSTS of biological motion or detection of intentional agents. [61]. These results suggest the existence of a functional gradient of social cognition, starting from the low-level detection

of other agents in the pSTS and EBA and ending in the more anterior and dorsal areas activated by ToM and VPT with the high-level mental representation of the other agent decoupled from oneself. Our proposal is in line with the nexus model of the TPJ, which proposes that the basic social perception begins in the lateral occipital cortex and then converges dorsally into the abstract representation for social cognition [59], together with functional heterogeneity in the TPJ [55]. The TPJ is also associated with other types of perspective taking of others under complex social interactions, including joint action [62], charitable giving [63], strategic competitive interaction [64], social judgment or decision making [52,65]. Further studies are needed to understand the relationships between these functions and clarify the neural mechanisms and functional gradients within the TPJ.

A recent popular fMRI analysis uses multi-voxel pattern analysis (MVPA) to investigate the differences in spatial patterns of the individual voxel rather than the averaged activation intensities [66]. Although MVPA allows successful decoding by utilizing detailed spatial activity patterns, it has the disadvantage of losing spatial specificity. By using peak coordinates of activation for each condition within subjects, this study could reveal the difference in spatial location for VPT and ToM within the TPJ.

In summary, we investigated whether VPT and ToM have the same neural substrates in the TPJ by directly comparing the activations of individual participants using within-subjects fMRI design. Our findings revealed that VPT and ToM have gradient representations, indicating the functional heterogeneity of social cognition within the TPJ.

Figures

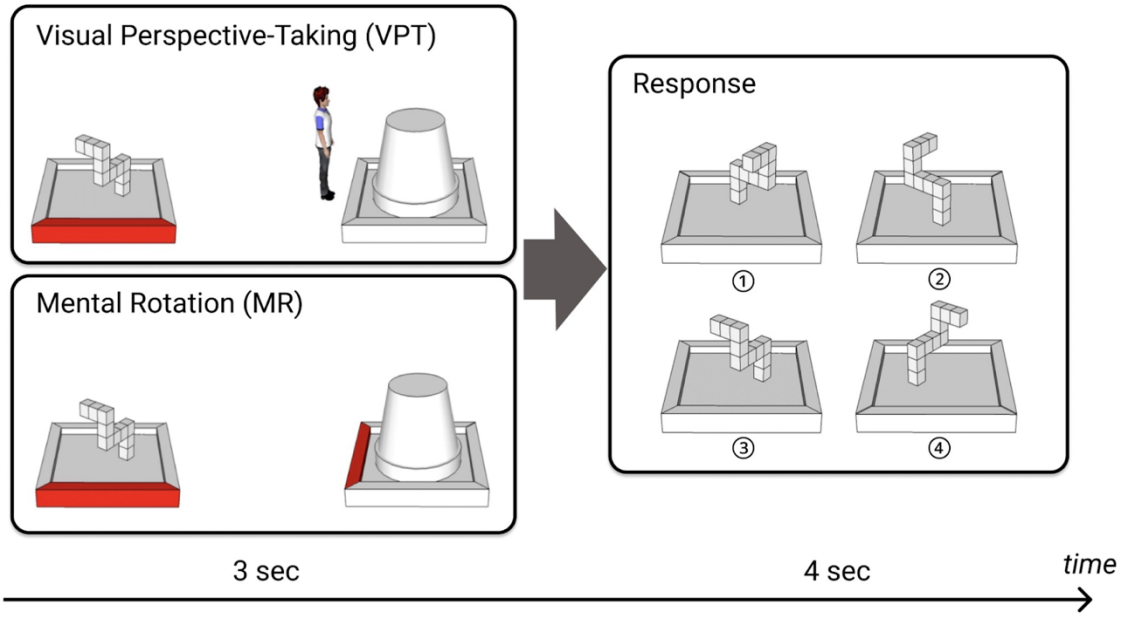

Fig. 1: Stimuli used for the VPT and MR tasks. First, a 3D object and a pot, which were placed on a tray with the front side colored red, were displayed for 3 s. In addition, an image of a person (avatar) standing either on the front, back, left, or right side of the pot was displayed in the VPT condition. The participants were asked to imagine the view of this object from the perspective of this person. In the MR condition, one of the four sides of the tray was colored in red. The participants were asked to imagine the rotation of the tray in the direction indicated by the red side of the tray as well as the rotated view of the object. In both VPT and MR conditions, four different views of the same object were then presented, and the participants had 4 s to select the correct response. The correct response depends on the condition; No. 1 for VPT, and No. 2 for MR in this illustration.

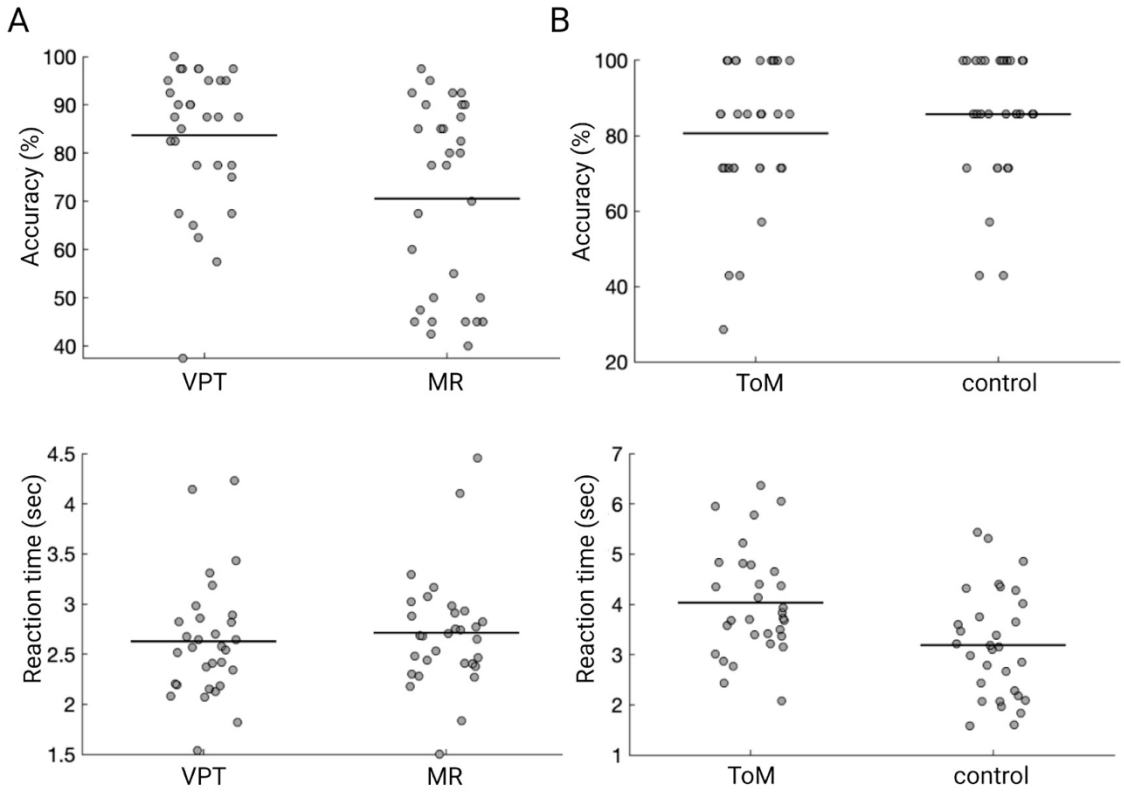

Fig. 2: Behavioral results for each run. (A) Task accuracy and reaction time (RT) in the VPT and MR conditions. (B) Task accuracy and RT in the ToM and control conditions. Each dot represents the data of an individual subject. The horizontal bar indicates the mean value. The individual data are available at https://doi.org/10.6084/m9.figshare.17149094

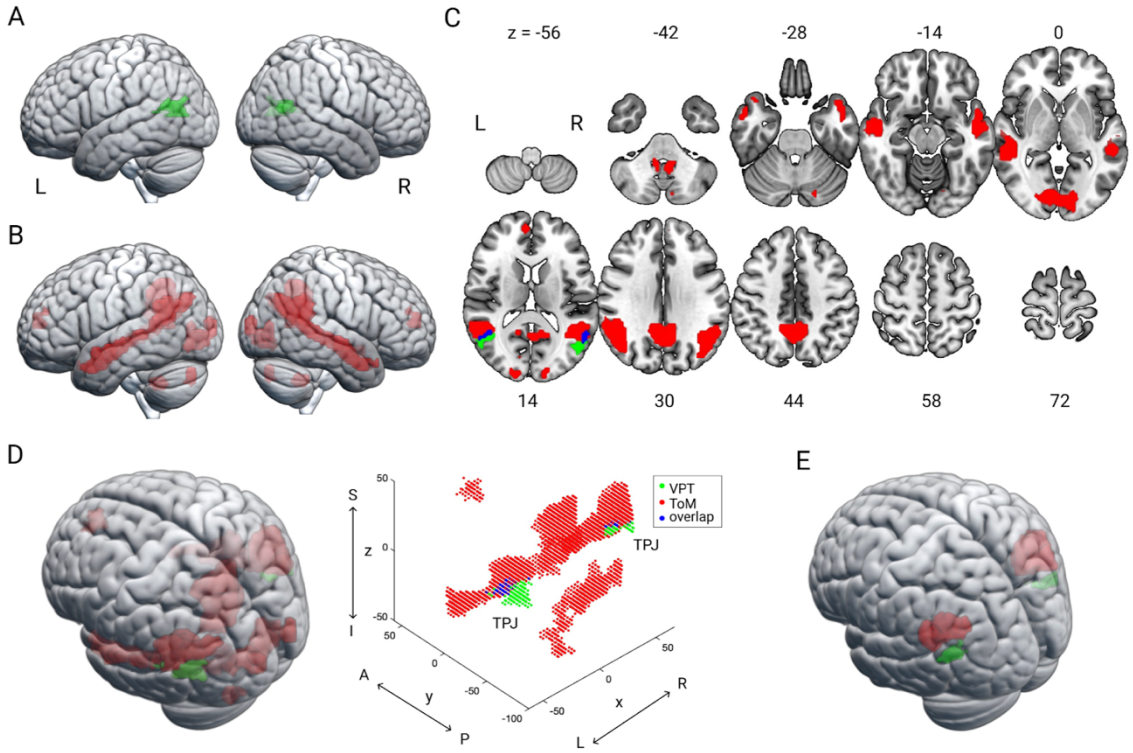

Fig. 3: The activated regions on the surface and in horizontal planes displayed with MRIcroGL. (A) Regions with significantly higher activation during the VPT task than during the MR task. (B) Regions with significantly higher activation during the ToM task than during the control trials. (C) The activated voxels obtained during the VPT (A in green, number of voxels = 130) and ToM (B in red, number of voxels = 3,124) tasks overlapped onto the same horizontal slices. The blue voxels are the voxels overlapping between the VPT and ToM conditions (number of voxels = 69). (D) The activated voxels obtained during VPT and ToM were overlaid onto the 3D template brain (left) and displayed with the 3D plot (right) in the same orientation. (E) The activated regions by direct comparison of the group-level activation maps between VPT (green) and ToM (red). . L, left; R, right; S, superior; I, inferior; A, anterior; P, posterior. The activated areas are listed in Table 1. The unthresholded activation images in NIfTY format are available at https://doi.org/10.6084/m9.figshare.17046179

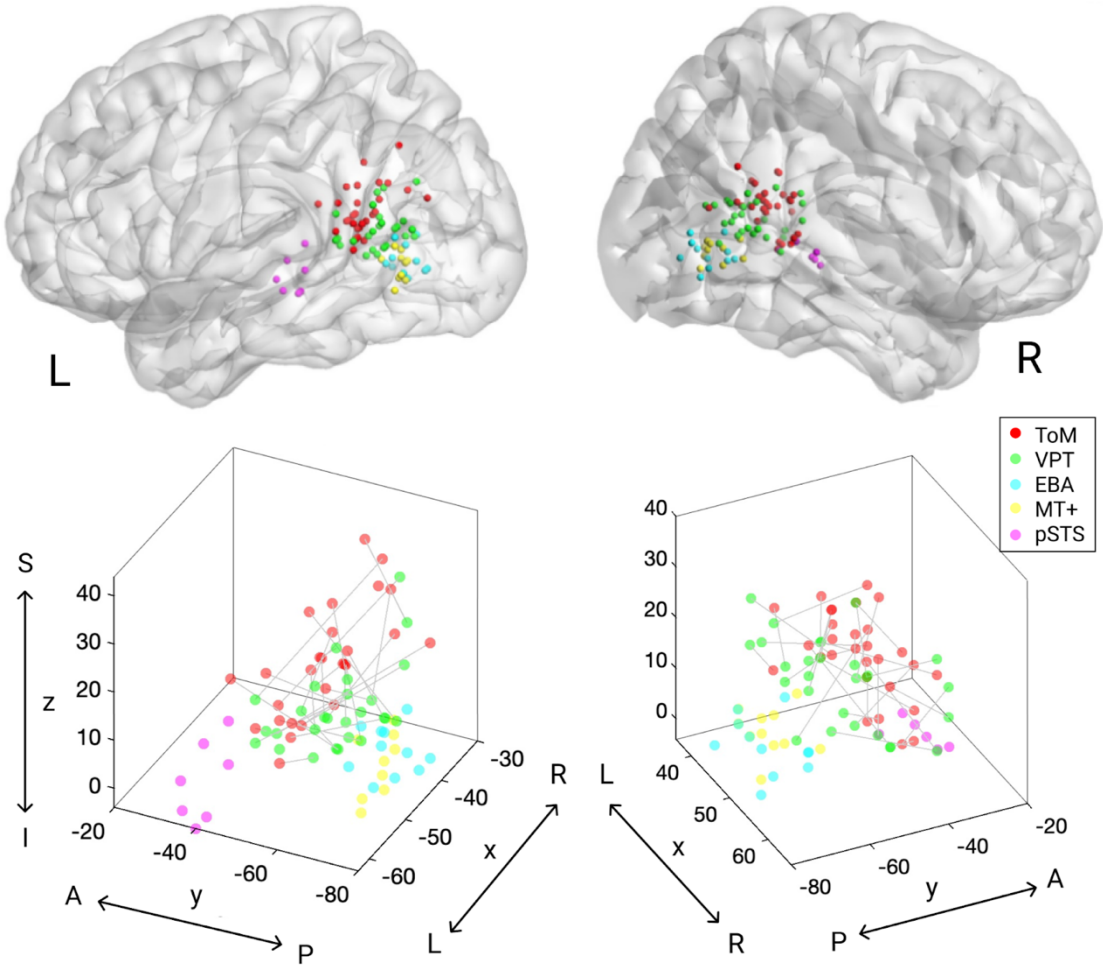

Fig. 4: Individual peak voxels obtained during VPT (green) and ToM (red) task as well as for EBA (cyan), MT+ (yellow), and pSTS (magenta) displayed on the glass brain using BrainNet Viewer (Xia et al. 2013) and the 3D plot on the MNI coordinates (lower panel). L, left; R, right; S, superior; I, inferior; A, anterior; P, posterior. The individual coordinate values are available at https://doi.org/10.6084/m9.figshare.17141867

Tables

Table 1: Anatomical regions, peak voxel coordinates, and t-values of observed activations.

Anatomic region	voxels	MNI coordinates			*t*-value
		x	y	z	
VPT > MR					
L TPJ	130	-51	-67	14	5.52
R TPJ	69	54	-64	14	4.78
ToM > ToM control					
R middle temporal cortex	1014	54	5	-28	9.15
R TPJ		48	-64	26	8.44
L middle temporal cortex	771	-54	-7	-16	8.29
L TPJ		-45	-58	23	7.00
M precuneus	787	3	-52	35	7.72
R lingual gyrus	443	6	-82	-4	7.23
L lingual gyrus		-9	-85	-7	5.57
L cerebellum	45	-21	-79	-34	6.02
M anterior cingulate cortex	59	6	53	11	5.58
L medial prefrontal cortex		-9	47	23	5.33
R medial prefrontal cortex		6	56	20	4.76
L cerebellum	74	-6	-55	-40	5.25
R cerebellum		9	-46	-40	4.80

(*VPT* > *MR*) > (*ToM* > *ToM control*)					
L TPJ	61	-48	-67	8	4.14
R TPJ	41	51	-70	8	3.03
(*ToM* > *ToM control*) > (*VPT* > *MR*)					
R TPJ	483	57	-49	20	4.65
L TPJ	300	-42	-55	23	4.41

MNI, Montreal Neurological Institute; L, left hemisphere; R, right hemisphere; M, medial; TPJ, temporo-parietal junction.

References

[1] A. Gunia, S. Moraresku, K. Vlček, Brain mechanisms of visuospatial perspective-taking in relation to object mental rotation and the theory of mind, Behav. Brain Res. 407 (2021) 113247.

[2] J.H. Flavell, The development of knowledge about visual perception, Nebr. Symp. Motiv. 25 (1977) 43–76.

[3] A. Pearson, D. Ropar, A.F. de C. Hamilton, A review of visual perspective taking in autism spectrum disorder, Front. Hum. Neurosci. 7 (2013) 1–10.

[4] S.M. Gzesh, C.F. Surber, Visual perspective-taking skills in children, Child Dev. 56 (1985) 1204–1213.

[5] J.H. Flavell, B.A. Everett, K. Croft, E.R. Flavell, Young children's knowledge about visual perception: Further evidence for the Level 1–Level 2 distinction, Dev. Psychol. 17 (1981) 99–103.

[6] P. Michelon, J.M. Zacks, Two kinds of visual perspective taking, Percept. Psychophys. 68 (2006) 327–337.

[7] K. Kessler, H. Rutherford, The two forms of visuo-spatial perspective taking are differently embodied and subserve different spatial prepositions, Front. Psychol. 1 (2010) 1–12.

[8] M. Hirai, Y. Muramatsu, M. Nakamura, Role of the Embodied Cognition Process in Perspective-Taking Ability During Childhood, Child Dev. (2018). https://doi.org/10.1111/cdev.13172.

[9] C.D. Frith, U. Frith, Interacting minds--a biological basis, Science. 286 (1999) 1692–1695.

[10] A.F. de C. Hamilton, R. Brindley, U. Frith, Visual perspective taking impairment in children with autistic spectrum disorder, Cognition. 113 (2009) 37–44.

[11] D. Samson, I.A. Apperly, C. Chiavarino, G.W. Humphreys, Left temporoparietal junction is necessary for representing someone else's belief, Nat. Neurosci. 7 (2004) 499–500.

[12] I.A. Apperly, D. Samson, C. Chiavarino, G.W. Humphreys, Frontal and temporo-parietal lobe contributions to theory of mind: neuropsychological evidence from a false-belief task with reduced language and executive demands, J. Cogn. Neurosci. 16 (2004) 1773–1784.

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

[14] I. Santiesteban, M.J. Banissy, C. Catmur, G. Bird, Functional lateralization of temporoparietal junction - imitation inhibition, visual perspective-taking and theory of mind, Eur. J. Neurosci. 42 (2015) 2527–2533.

[15] A.K. Martin, K. Kessler, S. Cooke, J. Huang, M. Meinzer, The Right Temporoparietal Junction Is Causally Associated with Embodied Perspective-taking, Journal of Neuroscience. 40 (2020) 3089–3095.

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

[17] M. Schurz, M. Kronbichler, S. Weissengruber, A. Surtees, D. Samson, J. Perner, Clarifying the role of theory of mind areas during visual perspective taking: Issues of spontaneity and domain-specificity, Neuroimage. 117 (2015) 386–396.

[18] S.M. Agarwal, V. Shivakumar, S.V. Kalmady, V. Danivas, A.C. Amaresha, A. Bose, J.C. Narayanaswamy, M.A. Amorim, G. Venkatasubramanian, Neural correlates of a perspective-taking task using in a realistic three-dimmensional environment based task: A pilot functional magnetic resonance imaging study, Clin. Psychopharmacol. Neurosci. 15 (2017) 276–280.

[19] C. Corradi-Dell'Acqua, K. Ueno, A. Ogawa, K. Cheng, R.I. Rumiati, A. Iriki, Effects of shifting perspective of the self: An fMRI study, Neuroimage. 40 (2008) 1902–1911.

[20] D. Dodell-Feder, J. Koster-Hale, M. Bedny, R. Saxe, fMRI item analysis in a theory of mind task, Neuroimage. 55 (2011) 705–712.

[21] Y. Otsuka, N. Osaka, T. Ikeda, M. Osaka, Individual differences in the theory of mind and superior temporal sulcus, Neurosci. Lett. 463 (2009) 150–153.
[22] M. Schurz, J. Radua, M. Aichhorn, F. Richlan, J. Perner, Fractionating theory of mind: a meta-analysis of functional brain imaging studies, Neurosci. Biobehav. Rev. 42 (2014) 9–34.
[23] R. Saxe, N. Kanwisher, People thinking about thinking peopleThe role of the temporo-parietal junction in "theory of mind," Neuroimage. 19 (2003) 1835–1842.
[24] P. Kanske, A. Böckler, F.M. Trautwein, T. Singer, Dissecting the social brain: Introducing the EmpaToM to reveal distinct neural networks and brain-behavior relations for empathy and Theory of Mind, Neuroimage. 122 (2015) 6–19.
[25] M. Schurz, M. Aichhorn, A. Martin, J. Perner, Common brain areas engaged in false belief reasoning and visual perspective taking: a meta-analysis of functional brain imaging studies, Front. Hum. Neurosci. 7 (2013) 712.
[26] A. Arora, M. Schurz, J. Perner, Systematic Comparison of Brain Imaging Meta-Analyses of ToM with vPT, Biomed Res. Int. 2017 (2017) 1–12.
[27] Y.-W. Hong, Y. Yoo, J. Han, T.D. Wager, C.-W. Woo, False-positive neuroimaging: Undisclosed flexibility in testing spatial hypotheses allows presenting anything as a replicated finding, Neuroimage. 195 (2019) 384–395.
[28] R.B. Mars, J. Sallet, U. Schüffelgen, S. Jbabdi, I. Toni, M.F.S. Rushworth, Connectivity-based subdivisions of the human right "temporoparietal junction area": evidence for different areas participating in different cortical networks, Cereb. Cortex. 22 (2012) 1894–1903.
[29] J. Decety, C. Lamm, The role of the right temporoparietal junction in social interaction: How low-level computational processes contribute to meta-cognition, Neuroscientist. 13 (2007) 580–593.
[30] D. Bzdok, R. Langner, L. Schilbach, O. Jakobs, C. Roski, S. Caspers, A.R. Laird, P.T. Fox, K. Zilles, S.B. Eickhoff, Characterization of the temporo-parietal junction by combining data-driven parcellation, complementary connectivity analyses, and functional decoding, Neuroimage. 81 (2013) 381–392.
[31] L. Young, D. Dodell-Feder, R. Saxe, What gets the attention of the temporo-parietal junction? An fMRI investigation of attention and theory of mind, Neuropsychologia. 48 (2010) 2658–2664.
[32] R.C. Oldfield, The assessment and analysis of handedness: the Edinburgh inventory, Neuropsychologia. 9 (1971) 97–113.
[33] T. Hatta, Z. Nakatsuka, Handedness inventory, in: Papers on Celebrating 63rd Birthday of Prof. Ohnishi, Osaka City University, 1975: pp. 224–245.
[34] R.N. Shepard, J. Metzler, Mental rotation of three-dimensional objects, Science. 171 (1971) 701–703.
[35] A. Ogawa, R. Yokoyama, T. Kameda, Development of a Japanese version of a theory-of-mind functional localizer for functional magnetic resonance imaging, The Japanese Journal of Psychology. 88 (2017) 366–375.
[36] D. Alcalá-López, J. Smallwood, E. Jefferies, F. Van Overwalle, K. Vogeley, R.B. Mars, B.I. Turetsky, A.R. Laird, P.T. Fox, S.B. Eickhoff, D. Bzdok, Computing the Social Brain Connectome Across Systems and States, Cereb. Cortex. 28 (2018) 2207–2232.
[37] P.E. Downing, Y. Jiang, M. Shuman, N. Kanwisher, A cortical area selective for visual processing of the human body, Science. 293 (2001) 2470–2473.
[38] M. Spiridon, B. Fischl, N. Kanwisher, Location and spatial profile of category-specific regions in human extrastriate cortex, Hum. Brain Mapp. 27 (2006) 77–89.
[39] M. Wilms, S.B. Eickhoff, K. Specht, K. Amunts, N.J. Shah, A. Malikovic, G.R. Fink, Human V5/MT+: Comparison of functional and cytoarchitectonic data, Anat. Embryol. . 210 (2005) 485–495.

[40] R.B. Tootell, J.B. Reppas, K.K. Kwong, R. Malach, R.T. Born, T.J. Brady, B.R. Rosen, J.W. Belliveau, Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging, J. Neurosci. 15 (1995) 3215–3230.
[41] E. Grossman, M. Donnelly, R. Price, D. Pickens, V. Morgan, G. Neighbor, R. Blake, Brain areas involved in perception of biological motion, J. Cogn. Neurosci. 12 (2000) 711–720.
[42] H. Peuskens, J. Vanrie, K. Verfaillie, G.A. Orban, Specificity of regions processing biological motion, Eur. J. Neurosci. 21 (2005) 2864–2875.
[43] Y. Okamoto, H. Kosaka, R. Kitada, A. Seki, H.C. Tanabe, M.J. Hayashi, T. Kochiyama, D.N. Saito, H.T. Yanaka, T. Munesue, M. Ishitobi, M. Omori, Y. Wada, H. Okazawa, T. Koeda, N. Sadato, Age-dependent atypicalities in body- and face-sensitive activation of the EBA and FFA in individuals with ASD, Neurosci. Res. 119 (2017) 38–52.
[44] M. Xia, J. Wang, Y. He, BrainNet Viewer: a network visualization tool for human brain connectomics, PLoS One. 8 (2013) e68910.
[45] M. Corbetta, G. Patel, G.L. Shulman, The reorienting system of the human brain: from environment to theory of mind, Neuron. 58 (2008) 306–324.
[46] J.P. Mitchell, Activity in right temporo-parietal junction is not selective for theory-of-mind, Cereb. Cortex. 18 (2008) 262–271.
[47] J.J. Geng, S. Vossel, Re-evaluating the role of TPJ in attentional control: contextual updating?, Neurosci. Biobehav. Rev. 37 (2013) 2608–2620.
[48] I. Santiesteban, S. Kaur, G. Bird, C. Catmur, Attentional processes, not implicit mentalizing, mediate performance in a perspective-taking task: Evidence from stimulation of the temporoparietal junction, Neuroimage. 155 (2017) 305–311.
[49] S.C. Krall, C. Rottschy, E. Oberwelland, D. Bzdok, P.T. Fox, S.B. Eickhoff, G.R. Fink, K. Konrad, The role of the right temporoparietal junction in attention and social interaction as revealed by ALE meta-analysis, Brain Struct. Funct. 220 (2015) 587–604.
[50] J. Scholz, C. Triantafyllou, S. Whitfield-Gabrieli, E.N. Brown, R. Saxe, Distinct regions of right temporo-parietal junction are selective for theory of mind and exogenous attention, PLoS One. 4 (2009) e4869.
[51] J.M. Zacks, J.M. Vettel, P. Michelon, Imagined viewer and object rotations dissociated with event-related FMRI, J. Cogn. Neurosci. 15 (2003) 1002–1018.
[52] L. Young, J.A. Camprodon, M. Hauser, A. Pascual-Leone, R. Saxe, Disruption of the right temporoparietal junction with transcranial magnetic stimulation reduces the role of beliefs in moral judgments, Proc. Natl. Acad. Sci. U. S. A. 107 (2010) 6753–6758.
[53] M. Sommer, K. Döhnel, B. Sodian, J. Meinhardt, C. Thoermer, G. Hajak, Neural correlates of true and false belief reasoning, Neuroimage. 35 (2007) 1378–1384.
[54] C. Kobayashi, G.H. Glover, E. Temple, Children's and adults' neural bases of verbal and nonverbal "theory of mind," Neuropsychologia. 45 (2007) 1522–1532.
[55] S.M. Lee, G. McCarthy, Functional Heterogeneity and Convergence in the Right Temporoparietal Junction, Cereb. Cortex. 26 (2016) 1108–1116.
[56] M. Aichhorn, J. Perner, B. Weiss, M. Kronbichler, W. Staffen, G. Ladurner, Temporo-parietal junction activity in theory-of-mind tasks: Falseness, beliefs, or attention, J. Cogn. Neurosci. 21 (2009) 1179–1192.
[57] D.M. Amodio, C.D. Frith, Meeting of minds: the medial frontal cortex and social cognition, Nat. Rev. Neurosci. 7 (2006) 268–277.
[58] C.M. Heyes, C.D. Frith, The cultural evolution of mind reading, Science. 344 (2014) 1243091.
[59] R.M. Carter, S.A. Huettel, A nexus model of the temporal – parietal junction, Trends Cogn. Sci. 17 (2013) 328–336.
[60] B. Kubit, A.I. Jack, Rethinking the role of the rTPJ in attention and social cognition in light of the opposing domains hypothesis: findings from an ALE-based meta-analysis and resting-state functional connectivity, Front. Hum. Neurosci. 7 (2013) 323.

[61] B. Chakrabarti, S. Baron-Cohen, Empathizing: neurocognitive developmental mechanisms and individual differences, Prog. Brain Res. 156 (2006) 403–417.
[62] M.O. Abe, T. Koike, S. Okazaki, S.K. Sugawara, K. Takahashi, K. Watanabe, N. Sadato, Neural correlates of online cooperation during joint force production, Neuroimage. 191 (2019) 150–161.
[63] A. Tusche, A. Böckler, P. Kanske, F.M. Trautwein, T. Singer, Decoding the charitable brain: Empathy, perspective taking, and attention shifts differentially predict altruistic giving, Journal of Neuroscience. 36 (2016) 4719–4732.
[64] A. Ogawa, T. Kameda, Dissociable roles of left and right temporoparietal junction in strategic competitive interaction, Soc. Cogn. Affect. Neurosci. 14 (2020) 1037–1048.
[65] T. Kameda, K. Inukai, S. Higuchi, A. Ogawa, H. Kim, T. Matsuda, M. Sakagami, Rawlsian maximin rule operates as a common cognitive anchor in distributive justice and risky decisions, Proc. Natl. Acad. Sci. U. S. A. 113 (2016) 11817–11822.
[66] K.A. Norman, S.M. Polyn, G.J. Detre, J.V. Haxby, Beyond mind-reading: multi-voxel pattern analysis of fMRI data, Trends Cogn. Sci. 10 (2006) 424–430.