NeuroImage 203 (2019) 116193

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Defining the neural correlates of spontaneous theory of mind (ToM): An fMRI multi-study investigation

Sara Boccadoro[a], Emiel Cracco[b], Anna R. Hudson[a], Lara Bardi[b,c], Annabel D. Nijhof[a,d], Jan R. Wiersema[a], Marcel Brass[b], Sven C. Mueller[a,*]

[a] *Department of Experimental Clinical and Health Psychology, Ghent University, Ghent, Belgium*
[b] *Department of Experimental Psychology, Ghent University, Ghent, Belgium*
[c] *Institute of Cognitive Neuroscience Marc Jeannerod, CNRS / UMR 5229, 67 Bd Pinel, 69500, Bron, France*
[d] *Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, UK*

ARTICLE INFO

Keywords:
Theory of mind
fMRI
Spontaneous mentalizing
False belief
Temporo-parietal junction

ABSTRACT

There is a major debate in the theory of mind (ToM) field, concerning whether spontaneous and explicit ToM are based on the same or two distinct cognitive systems. While extensive research on the neural correlates of explicit ToM has demonstrated involvement of the temporo-parietal junction (TPJ) and the medial prefrontal cortex (mPFC), few studies investigated spontaneous ToM, leaving some open questions. Here, we implemented a multi-study approach by pooling data from three fMRI studies to obtain a larger sample to increase power and sensitivity to better define the neurocognitive mechanisms underlying spontaneous ToM. Participants watched videos in which an agent acquires a true or false belief about the location of a ball. Thus, the belief of the agent and that of the participant could either match or differ. Importantly, participants were never asked to consider the belief of the agent and were only instructed to press a button when they detected the presence of the ball after an occluder fell at the end of each video. By analysing the blood-oxygen level dependent signal during the belief formation phase for false versus true beliefs, we found a cluster of activation in the right, and to a lesser extent, left posterior parietal cortex spanning the TPJ, but no mPFC activation. Region of interest (ROI) analysis on bilateral TPJ and mPFC confirmed these results and added evidence to the asymmetry in laterality of the TPJ in spontaneous ToM. Interestingly, the whole brain analysis, supported by an overlap with brain maps, revealed maximum activation in areas involved in visuospatial working memory and attention switching functions, such as the supramarginal gyrus, the middle temporal gyrus, and the inferior frontal gyrus. By contrast, evidence for the presence of brain-behaviour correlations was mixed and there was no evidence for functional connectivity between the TPJ and mPFC. Taken together, these findings help clarifying the brain system supporting spontaneous ToM.

1. Introduction

Humans are a social species and so, by definition, regularly engage in social interactions, which require the ability to understand and predict the goals, beliefs, desires, thoughts, and behaviours of other people. This fundamental ability is called Theory of Mind (ToM) and is crucial for representing others' mental states. Traditionally, ToM has been investigated using false belief tasks, mainly the 'Sally-Anne' false belief task (Baron-Cohen et al., 1985; Wimmer and Perner, 1983), which *explicitly* asks participants to reason about other people's mental states. In the Sally-Anne paradigm, participants watch a scene in which a character, Sally, places an object in a box before leaving the scene. After Sally leaves, another character Anne moves the object to a different box. When Sally re-enters the scene, participants are asked to indicate in which box they think Sally will look for the object. Indicating the correct box requires the capacity to represent Sally's false belief. This line of research suggests that ToM requires the executive function ability to suppress one's own belief, which emerges later in development and, thus, only children aged four years old and above tend to pass the classic false belief task (Wellman et al., 2001; Wimmer and Perner, 1983). However, subsequent research has challenged this idea, showing that infants before this age, much like adults, can already represent others' belief in a *spontaneous* way, when they are not explicitly required to do so or when others' mental states are irrelevant for their goals (Clements and Perner,

[*] Corresponding author. Henri Dunantlaan 2, 9000, Ghent, Belgium.
E-mail address: muellersvenc@gmail.com (S.C. Mueller).

https://doi.org/10.1016/j.neuroimage.2019.116193
Received 4 March 2019; Received in revised form 12 August 2019; Accepted 13 September 2019
Available online 13 September 2019

1994; Kovács et al., 2010; Onishi and Baillargeon, 2005; Schneider et al., 2017; Southgate et al., 2007).

This spontaneous ToM ability can be investigated with tasks in which participants are not specifically instructed to think about others' mental states nor are explicitly asked to report any information regarding mental states. Thus, in spontaneous ToM tasks, participants spontaneously track the belief of an agent. For example, Kovács et al. (2010) developed a novel behavioural object detection task, where participants watched a video depicting an agent (a smurf) acquiring information about the location of an object (a ball) that could be behind an occluder or not. Here, the belief of the agent and that of the participant could either match, in the True Belief condition (the participant and the agent believe the object to be at the same location), or differ, in the False Belief condition (the participant and agent believe the object to be at different locations). Participants were never asked to consider the belief of the agent and were only asked to press a button when the ball was present after the occluder fell. Importantly, whether the ball is present or absent is completely random in this task. Nevertheless, participants were biased by the beliefs of the agent, resulting in faster responses when the agent believed the ball was present, even when they themselves knew the ball should not be present. This pattern of behavioural responses suggests that participants spontaneously engaged in ToM.

Mounting work has documented the neural network involved in explicit ToM, revealing consistent involvement of a pattern of brain regions forming the so-called "ToM network", including the temporoparietal junction (TPJ), the superior temporal sulcus (STS), the precuneus (PC), the temporal poles and the medial prefrontal cortex (mPFC) (Fletcher et al., 1995; Gallagher et al., 2000; Ruby and Decety, 2003; Saxe and Kanwisher, 2003). This is further cemented by meta-analytic evidence of explicit ToM data suggesting consistent activation of the TPJ and the mPFC (Decety and Lamm, 2007; Molenberghs et al., 2016; Schurz et al., 2014; Van Overwalle, 2009). By contrast, neural findings on spontaneous ToM tasks are presently mixed, likely due to fewer available studies with small samples (Bardi et al., 2017 (N = 22); Kovács et al. 2014 (N = 15); Naughtin et al., 2017 (N = 22); Schneider et al., 2014 (N = 16)). For example, Kovács et al. (2014) and Bardi et al. (2017) reported enhanced TPJ activation for false relative to true beliefs in the belief formation phase. However, contrary to explicit ToM studies, there was no mPFC activation during this phase. Moreover, while TPJ activation was missing altogether in a spontaneous ToM task by Schneider et al. (2014), Naughtin et al. (2017) reported higher activity in a more extensive network comprising the right TPJ, precuneus, left MFG and right STS for false belief trials relative to no belief trials. Therefore, before direct comparison to explicit ToM tasks can be made, more evidence regarding the neural network underlying spontaneous ToM is needed.

A related contentious issue is whether the observed TPJ activity is stronger on the right (vs. the left) side. While most ToM studies generally report larger activation of the right vs. the left TPJ, some studies claim that the right TPJ is more specifically involved in explicit ToM than its left counterpart (Aichhorn et al., 2009; Döhnel et al., 2012; Liu et al., 2009; Saxe, 2010), and yet others favour a more bilateral activation of the TPJ in ToM tasks (Jenkins and Mitchell, 2010; Krall et al., 2015; Saxe and Kanwisher, 2003). More recently, Kovács et al. (2014), Bardi et al. (2017) and Nijhof et al. (2018) reported the recruitment of only the right TPJ in spontaneous ToM, but also left open the question about a possible role of the left TPJ, given the association between damage to the left TPJ and explicit ToM deficits as reported in lesion studies (Apperly et al., 2004; Samson et al., 2004). In other words, previous studies have not directly tested the hypothesis that spontaneous ToM is right (versus left) lateralized.

As noted above, some of the inconsistencies emerging between the few available spontaneous ToM studies may be due to the subtle differences between the paradigms used, or to the limited number of participants recruited. Indeed, recent efforts have been launched to deal with the replication crisis in psychology (Maxwell et al., 2015) and the neurosciences (Poldrack et al., 2017). With regards to the present topic, the majority of theory of mind studies have samples sizes of <30 participants (for overviews see Molenberghs et al., 2016; Schurz et al., 2014) with few notable exceptions (Durfour et al., 2013; Moessnang et al., 2016). Samples sizes of $n < 30$ are common in group fMRI research but have been deemed insufficient (Poldrack et al., 2017). This has direct implications for the power and sensitivity of these studies. Geuter et al. (2018) recently investigated the effects of sample sizes on the sensitivity and reliability of group fMRI analysis, demonstrating that sample sizes of 40 are able to identify regions with large effect sizes (Cohen's $d > 0.8$), while sample sizes closer to 80 can detect regions with medium effect sizes ($0.5 < d < 0.8$). In fact, an examination of the average effect sizes for fMRI studies by Poldrack et al. (2017) indicated that most were in the small to medium effect range (ranging between $d = 0.4 - d = 0.6$), suggesting that larger sample sizes are needed in ToM research to reliably detect realistic effect sizes. The current sample size of 68 participants provides us with the necessary power to detect medium effect sizes in mentalizing studies and, thus, with more power to answer novel questions, some of which could not be answered in previous research. For example, specifically in relation to brain-behaviour correlations, the Poldrack et al. (2017) study also included a simulation that showed that spurious brain behaviour correlations, a third contentious issue in spontaneous ToM research, may emerge with sample sizes of less than 30 participants and might be unreliable. In relation to this, a recent ToM study on the reliability and sensitivity of brain activity during mentalizing (Moessnang et al., 2016, N = 46) indicated good to excellent reliability for bilateral posterior temporal sulcus but comparatively poor performance of the mPFC.

Therefore, given a) the scarcity of studies investigating the spontaneous ToM system, b) contradictory findings, and c) small sample sizes, the present study investigated, using a multi-study design in a sample of healthy participants, 5 main questions. First, it sought to uncover the potential involvement of the mPFC in the formation of belief, second, to examine the lateralization of TPJ activation, third, to examine the spontaneous ToM network within a broader context and relative to published socio-cognitive activations maps, fourth, to assess potential brain-behaviour correlations, and fifth, the functional connectivity between ToM regions. With regards to the latter, to our knowledge, functional connectivity during spontaneous ToM has scarcely been investigated. While Burnett and Blakemoore (2009) documented increased functional connectivity between the mPFC and the TPJ, Moessnang et al. (2017) reported such connectivity between the mPFC and posterior superior temporal sulcus, i.e., more ventral to the TPJ. The main predictions, in turn for each of the questions, were as follows: firstly, in addition to the already established activation of the TPJ, we expected to find activation of the mPFC, a region that has consistently been reported in the explicit ToM literature (cf. Schurz et al., 2014). Secondly, directly testing the issue of laterality, we predicted greater activity in the right TPJ than in the left TPJ, which would confirm the right (versus left) dominance for spontaneous ToM processing. Thirdly, to assess the consistency of our findings with the broader ToM literature, we compared our whole brain activation map to the average ToM map using Neurosynth (http://www.neurosynth.org/). Fourth, we anticipated a positive correlation between brain activity in the belief formation phase and behaviour in the outcome phase. Finally, we aimed to replicate the previously reported increased functional connectivity between temporo-parietal regions and the mPFC (Burnett and Blakemoore, 2009; Moessnang et al., 2017).

2. Materials and methods

2.1. Participants

This multi-study includes samples from three different, independent studies conducted at Ghent University (Bardi et al., 2017, Hudson et al., [preprint, bioRxiv], Nijhof et al., 2018), all of which used the same identical spontaneous ToM task but with different participants. In total, data from 74 healthy participants were available. Six subjects had to be excluded due to excessive movement (>3 mm or 3° on any dimension), resulting in a total final sample of 68 participants (17 males; mean age = 31.13 years, SD = 10.49). All participants had normal or corrected-to-normal vision, did not have any reported history of neurological disorders and gave written informed consent prior to the study. Handedness information was not available for all participants included in this study. All studies were approved by the Ethical Committee of Ghent University Hospital.

2.2. Task and stimuli

In all three studies, participants performed a spontaneous ToM task, called the "Buzz Lightyear" task (Bardi et al., 2017; Nijhof et al., 2018), which is an adaptation of the task originally developed by Kovács et al. (2010). Participants laid in the MRI scanner while watching short videos and detecting an object at the end of each video. All movies consisted of two phases: the belief formation phase and the outcome phase. Each movie lasted 13.8 s. All movies started with the belief formation phase, in which an agent (*Buzz Lightyear*) placed a ball on a table in front of an occluder. The ball rolled behind the occluder at 3 s and after this, the movie could continue in four possible ways:

1. In the True Belief-Positive Content condition, the ball rolled out of the scene from behind the occluder and then rolled back behind it at 10 s in the presence of the agent, who then left the scene at 11 s. As a consequence, both the participant (P) and the agent (A) believed that the ball was behind the occluder (P + A+).
2. In the True Belief-Negative Content condition, after emerging from behind the occluder without leaving the scene, the ball rolled back behind the occluder and then left the scene at 10 s in the presence of the agent, who left the scene at 11 s. Therefore, both the participant and the agent believed that the ball was not behind the occluder (P-A-).
3. In the False Belief-Positive Content condition, the ball was behind the occluder when the agent left the scene at 6 s. Then the ball emerged from behind the occluder without leaving the scene, rolled back behind the occluder and finally left the scene at 11 s, when the agent was absent. Thus, the participant believed that the ball was not behind the occluder, whereas the agent wrongly believed that the ball was behind the occluder (P-A+).
4. In the False Belief-Negative Content condition, the ball rolled out of the scene while the agent was present. The agent left the scene at 9 s and, in his absence, the ball rolled back behind the occluder at 11 s. Therefore, the participant believed that the ball was behind the occluder, whereas the agent wrongly believed that the ball was not behind the occluder (P + A-).

To ensure that attention was maintained throughout the presentation of the movies, participants had to press a button with their left index finger as quickly as possible when Buzz left the scene.

In the outcome phase, at the end of each movie, the agent re-entered the scene, the occluder fell down and there could be two possible and equally probable outcomes, in which the ball could be either present or absent behind the occluder. At this point, participants had to press a button with their right index finger as quickly as possible, but only if the ball was present after the occluder fell. The presence (B+) or absence (B-) of the ball was completely independent of the belief formation phase, since the ball was present randomly in half of the trials. Thus, the ball could be expected or unexpected for both the participant and the agent. The combination of belief formation phase (P-A-; P + A+; P + A-; P-A+) and outcome phase (B+ and B-) resulted in eight different movies. Each movie was repeated 10 times, thus the task consisted of 80 trials presented in a randomised order in two blocks (fMRI runs) of 40 trials, with a short break in between, except for Hudson's study, in which the 8 movies were repeated 8 times, thus resulting in 64 trials. The inter-trial interval was determined using a pseudo-logarithmic jitter with steps of 600 ms: half of the intervals were short (range from 200 to 2000 ms), one-third was intermediate (range from 2600 to 4400 ms) and one-sixth was long (range from 5000 to 6800 ms), with a mean inter-trial interval of 2700 ms. No instruction to reason about the agent's belief was given to participants. Two studies (Bardi et al., 2017; Nijhof et al., 2018) included also an explicit ToM version of the task, always performed *after* the two spontaneous runs. Only the two spontaneous ToM runs were included in the present study.

2.3. fMRI data acquisition

Structural T1-weighted MRI images were acquired using a 3T Siemens Magnetom TrioTim MRI scanner. More specifically, 176 slices of a T1-weighted MPRAGE high resolution structural image were acquired (repetition time (TR) = 2250 ms, echo time (TE) = 4.18 ms, image matrix = 256 × 256, field of view (FOV) = 256 mm, flip angle = 9°, slice thickness = 1 mm, voxel size = 1.0 × 1.0 x 1.0). The whole-brain T2*-weighted Echo Planar Images (EPI) sequence was identical across the three studies (TR = 2000 ms, TE = 28 ms, image matrix = 64 × 64, FOV = 224 mm, flip angle = 80°, slice thickness = 3.0 mm, voxel size = 3.5 × 3.5 × 3.0 mm, number of slices = 34). The number of volumes depends on how fast participants completed the task. The estimated number of volumes for two samples (Hudson et al., [preprint, bioRxiv]; Nijhof et al., 2018), was between 278 and 309, while for the other sample of participants (Bardi et al., 2017) it was between 373 and 417.

2.4. fMRI data pre-processing

For the sake of consistency across studies, all data were pre-processed again with SPM8 software (Wellcome Department of Cognitive Neurology, London, UK) in MATLAB (The Mathworks). The first four volumes for each EPI series were removed to allow magnetisation to reach a dynamic equilibrium. The pre-processing steps for the remaining volumes started with spatial realignment of the functional images using a rigid body transformation and then slice time correction of the realigned images with respect to the middle slice. The structural image of each subject was co-registered with the mean of the slice-time corrected images. During segmentation, the structural scans were brought in line with SPM8 tissue probability maps. The parameters estimated during the segmentation step were then used to normalise the functional images to standard MNI space. Lastly, the normalized functional images were resampled into 3 × 3 × 3 mm voxels and spatially smoothed with a Gaussian kernel of 8 mm (full-width at half maximum).

2.5. fMRI data analysis

2.5.1. Whole-brain analysis for spontaneous ToM network

Utilizing our larger sample size to a) examine the implicit ToM network across the brain and b) prepare the analysis for comparison with the broader literature and available brain maps (cf. section 2.5.4), first- and second-level analysis were carried out using SPM8 and the general linear model (GLM). Per run, the model contained four separate regressors for all combinations of Belief (true and false belief) and Belief Content (positive and negative content) in the belief formation phase, with durations of 9s modelled from the moment when the ball starts rolling on the table (3 s) to the moment when the agent re-enters the scene (12 s). In addition, to prevent the outcome phase signal from leaking into the error term, eight regressors were added to model the outcome phase, (all possible combinations of Belief, Belief Content and Outcome), with duration of 0 s, modelled at the point when the occluder has completely fallen down and the presence or absence of the ball is revealed. In total, there were 12 regressors of interest per run. Furthermore, six subject-specific movement regressors were added per run to account for head motion. All regressors were convolved with the canonical HRF.

The second-level analysis was conducted using a flexible factorial model with a between-subjects factor for experiment (to account for the fact that the data came from three different studies, which might lead to potential confounders such as stratification, experimenters, and time of testing), and a within-subjects factor for condition (P + A+, P-A-, P-A+, P + A-). In order to identify the regions involved in false belief tracking (belief formation phase), we computed our main contrast of interest (PxA, interaction contrast) as follows: False Belief (P-A+ and P + A-) > True Belief (P-A- and P + A+) and the reverse contrast for regions involved in true belief tracking. Additionally, we also computed the ToM Index (i.e., P-A+ > P-A-) to correlate it against the behavioural ToM index (cf. section 2.5.5). Results for these whole-brain analyses were corrected for multiple comparisons using a $p < 0.01$ FWE whole-brain corrected threshold and minimal extent of 20 contiguous voxels. Since our interest was to investigate the neural basis of false belief reasoning, we focused on the belief formation phase, in which the participant forms their own belief and that of the agent. Thus, the outcome phase was not taken into consideration for the analysis and no results are displayed for this phase.

To control for the influence of gender, an additional second-level analysis was conducted with this factor as a covariate of no interest.

2.5.2. *mPFC ROI analysis*

To assess the presence of the mPFC in spontaneous ToM, a ROI analysis based on the meta-analysis by Kovács et al. (2014) (N = 26 studies) was carried out, using a sphere with a 6 mm radius centred on the mPFC peak coordinate [MNI xyz: 2 53 13] reported by Kovács et al. (2014). Mean βs were extracted using the MARSBAR toolbox for SPM (Brett et al., 2002). The obtained β values were entered into a repeated-measure ANOVA containing Participant (P- or P+) and Agent (A- or A+) belief as within-subjects factors and experiment as a between-subjects factor. An alpha level of $p < 0.01$ was used. Furthermore, we also included a Bayes Factor (BF) analysis with default JASP priors to calculate the likelihood of the data under the alternative hypothesis (False Belief > True Belief) relative to the null hypothesis (False Belief = True Belief). For example, a BF of 3 means that the data are three times more likely under the alternative hypothesis than under the null hypothesis, while a BF of 0.33 means the opposite (Rouder et al., 2009).

2.5.3. *Lateralization of TPJ activity*

To assess TPJ lateralization, we again conducted an ROI analysis. That is, we first defined ROIs for the right and left TPJ by drawing spheres with 6 mm radii around the peak coordinates of the right [MNI xyz: 56–47 33] and left TPJ [-56 -47 33] obtained in the aforementioned meta-analysis by Kovács et al. (2014). We then carried out a repeated-measure ANOVA to explore the degree to which TPJ activation was lateralized, by entering the β values of left and right TPJ with Participant (P- or P+), Agent (A- or A+), and Location (left and right) as within-subjects factors and experiment as between-subjects factor. An alpha level of $p < 0.01$ was used.

2.5.4. *Cognitive decoding of brain activity: brain map comparison*

To estimate the consistency of our results with the average ToM activation found in the literature, the activation maps of both the False > True and True > False Belief PxA contrasts (cf. 2.5.1) were entered in Neurovault (https://neurovault.org/collections/VQVUVUMR/; Gorgolewski et al., 2015). Neurovault allows us to use the cognitive decoding feature to compare the uploaded maps with the activation maps associated with various cognitive functions across many papers, using spatial correlations calculated in Neurosynth (http://neurosynth.org/). As output, this analysis reveals the cognitive functions whose activation maps are most correlated with the uploaded maps. Among the first 15 most correlated functional terms, those belonging to the same functional category were taken (cf. Supplementary material). The identified functional categories for the False > True Belief contrast were working memory and visuospatial. The identified categories for the True > False Belief contrast were self-referential and emotion. An alpha level of $p < 0.01$ was used.

2.5.5. *Brain-behaviour relation*

Reaction times (RT) were recorded for the detection of the ball at the end of each movie. Specifically, the behavioural analysis focused on the difference in RTs between the P-A- and the P-A+ condition, known as the *ToM Index* (Deschrijver et al., 2016). This ToM Index is a measure of the influence of the agent's belief on ball detection, so that a larger ToM Index indicates a larger influence. This ToM index was calculated here to calculate the Pearson correlation of the neural ToM Index in the right and left TPJ ROIs with this behavioural ToM Index. The neural ToM Index was calculated by subtracting the (β) values for the P-A- condition to the (β) values for the P-A+ condition for each ROI.

2.5.6. *Functional connectivity (psychophysiological interaction, PPI)*

Given the near absence of functional connectivity analyses in prior spontaneous ToM work, we aimed to replicate the finding of increased functional connectivity between temporo-parietal regions and the mPFC (Burnett and Blakemoore, 2009; Moessnang et al., 2017). The (voxel-wise) gPPI analyses (McLaren et al., 2012) examined brain-wise functional connectivity of 1) the *a priori* rTPJ ROI and 2) of the two main temporo-parietal peaks that emerged from the whole-brain analysis (rMTG and SMG) with the rest of the brain for the False > True Belief contrast using 6 mm spheres around the peak coordinates ($p < 0.01$ FWE corrected, > 20 contiguous voxels).

3. Results

3.1. *Whole brain analysis*

Sampling the entire brain to identify the brain network involved in spontaneous ToM, the PxA False > True Belief contrast yielded activation in a large cluster spanning the temporo-parietal cortex including the TPJ. Closer inspection of this cluster revealed that two of the three strongest activation peaks were located in the temporo-parietal cortex, with one peak in the middle temporal gyrus (MTG) (MNI xyz: [60–52 1]) and another in the supramarginal gyrus (SMG) [54–37 46], while a third peak [12–73 4] belonged to a more occipital region (the lingual gyrus). In addition to this temporo-parietal cluster, the PxA False > True Belief contrast also revealed activity in the right inferior frontal gyrus (IFG), the left inferior parietal lobule (IPL), the left middle temporal gyrus (MTG) and the right middle frontal gyrus (MFG). The same pattern of activity was found for the ToM Index False > True Belief contrast, except that the lingual gyrus peak found in the PxA analysis was now replaced by a peak in the angular gyrus (AG) [39–46 37] (Table 1, Fig. 1, Fig. 2). The PxA True > False Belief contrast revealed activation in a single cluster in the medial orbitofrontal cortex. The same cluster was also found for the ToM Index True > False Belief contrast, with additional activation clusters in the occipital cortex and precentral gyrus.

3.2. *mPFC ROI analysis*

A main effect of the agent's belief emerged for the mPFC, $F(1,65) = 8.75$, $p = 0.004$, which was more activated when the agent believed the ball was absent (A-) than when he believed the ball was present (A+). However, neither the PxA interaction, $F(1,65) = 2.39$, $p = 0.127$ (Fig. 3), nor any of the other main effects or interactions were significant for the mPFC (Fig. 3). The absence of a PxA interaction was further supported by a Bayesian analysis, which revealed a BF of 0.052, indicating strong evidence against the hypothesis that the mPFC was more activated in the False compared to True Belief condition.

Table 1
Peaks of activation in the belief formation phase.

Area	MNI peak coordinates xyz	Cluster size	T
PxA			
False > True Belief			
rMTG	60 -52 1	4991	11.21
rLING	12 -73 4		11.13
rSMG	54 -37 46		11.03
rIFG	51 20 -8	903	8.56
	51 11 16		8.23
	57 17 1		8.00
lIPL	−51 -37 55	722	9.82
	−60 -46 37		7.25
lPoCG	−66 -19 22		7.16
rThalamus/rCAU	24 -28 10	355	7.55
	21 -13 16		6.96
	18 -1 19		6.64
lMTG/ITG	−54 -64 -2	170	6.83
lMTG/STG	−63 -52 16		5.08
rMFG	42 47 19	136	6.49
lCAU/lThalamus	−18 2 22	71	6.06
	−18 -16 -19		5.50
lInsula	−36 23 4	33	6.36
Vermis	3 -43 -2	21	6.29
True > False Belief			
lmOFC	−6 47 -11	62	6.24
ToM Index			
False > True Belief			
rMTG	60 -52 1	2552	10.17
rSMG	54 -37 46		9.92
rAnG	39 -46 37		9.17
rIFG	36 14 34	622	7.34
	51 14 19		7.05
rMFG	36 5 40		6.83
lIPL	−51 -37 55	135	7.61
	−30 -46 34		6.16
rLING	12 -73 4	127	7.30
rThalamus	21 -28 13	25	6.49
True > False Belief			
rIOG	33 -79 -11	384	8.83
rCAL	15 -97 4		8.65
rLING	24 -88 -8		8.21
lmOFC	−6 47 -8	243	6.98
lREC	−3 38 -14		6.84
rPreCG	36 -16 52	45	6.43
lIOG	−30 -94 -8	42	6.41
lCAL	−18 -103 -2		5.63
lMOG	−9 -103 4		5.52

MNI, Montral Neurological Institute; MTG, middle temporal gyrus; LING, lingual gyrus; SMG, supramarginal gyrus; AnG, angular gyrus; IFG, inferior frontal gyrus; IPL, inferior parietal lobule; PoCG, postcentral gyrus; CAU, caudate nucleus; MTG, middle temporal gyrus; MFG, middle frontal gyrus; mOFC, medial orbitofrontal cortex; IOG, inferior occipital gyrus; CAL, calcarine fissure; REC, gyrus rectus; PreCG, precentral gyrus; MOG, middle occipital gyrus; r, right; l, left; T, true; F, false.

3.3. Lateralization of TPJ activity in spontaneous ToM

The investigation of TPJ activity revealed a significant P × A interaction, $F(1,65) = 41.99$, $p < 0.001$, with stronger activation in the False Belief conditions (P-A+ and P + A-) than in the True Belief conditions (P-A- and P + A+). Furthermore, supporting lateralization, there was a significant location x P × A interaction, $F(1, 65) = 7.92$, $p = 0.006$, with a stronger P × A interaction for the right, $F(1, 65) = 44.18$, $p < 0.001$, than for the left, $F(1, 65) = 26.21$, $p < 0.001$, TPJ (Fig. 3), meaning that the difference in activation between the False and True Belief conditions was stronger in the right TPJ than in the left TPJ. These results confirm the asymmetry of TPJ activation. Importantly, no significant interactions with experiment emerged.

3.4. Cognitive decoding of brain activity

To evaluate the consistency of our task findings with the average pattern of activation associated with ToM in other tasks and designs, we compared the maps obtained in the whole-brain PxA analysis with the activation maps obtained from Neurosynth (see 2.5.4). Most interestingly, the activation map for the False > True Belief contrast corresponded more strongly to the activation maps related to working memory (all structures: $r = 0.23$; cortex: $r = 0.24$; subcortex: $r = 0.10$) and visuospatial functions (all structures: $r = 0.14$; cortex: $r = 0.15$; subcortex: $r = -0.01$) than to the map related to ToM (all structures: $r = -0.04$; cortex: $r = -0.06$; subcortex: $r = -0.07$), especially in the more dorsal active areas, according to visual inspection (Fig. 4, upper panel). The activation map for the True > False Belief contrast corresponded more strongly to the activation map related to self-referential processing (all structures: $r = 0.14$; cortex: $r = 0.18$; subcortex: $r = 0.01$) than to the map related to ToM (all structures: $r = 0.04$; cortex: $r = 0.06$; subcortex: $r = 0.07$) (Fig. 4, lower panel).

3.5. Brain-behaviour relation

Analysis of the ToM Index with a paired t-test revealed a significantly faster responding during the P-A+ relative to the P-A- condition, $t(65) = -4.57$, $p < 0.01$, confirming that participants responded faster when the agent believed the ball was present than when the agent believed the ball was absent. However, there was no significant correlation for either the rTPJ, $r = -0.206$, 95% CI: [-0.426 0.038], $p = 0.098$, or the lTPJ, $r = -0.237$ [-0.453 0.005], $p = 0.055$, between the behavioural and neural ToM index. Moreover, even though both correlations might have still been considered marginally significant at a more conventional $p < 0.05$ level, this was no longer the case after outlier removal, rTPJ: $r = -0.119$ [-0.356 0.133], $p = 0.352$; lTPJ: $r = -0.160$ [-0.391 0.089], $p = 0.206$.

3.6. Functional connectivity (psychophysiological interaction)

The PPI analyses did not reveal any significant findings, thus failing to characterize the functional network of spontaneous false belief processing. Even though no significant result emerged, the PPI analysis is reported here in-line with good practice standards of reporting negative findings.

4. Discussion

This study aimed to clarify the neural mechanisms of spontaneous ToM by pooling data from three studies in a multi-study fashion and to resolve some of the present controversies. With regards to hypothesis 1, while there was no effect of the mPFC, a large cluster in the temporo-parietal cortex including the TPJ emerged. ROI analyses confirmed these results, showing significant activation of bilateral TPJ and strong evidence against the involvement of the mPFC during the belief formation phase. Thus, the spontaneous ToM system appears to recruit the TPJ but not the mPFC. Secondly, ROI analyses confirmed that both the right and left TPJ were activated by inconsistencies between the belief of the participant and the belief of the agent but favouring the right side. Third, the comparison with Neurosynth maps revealed that our maps of activation overlapped more with maps related to other functions (working memory and visuospatial for the False > True Belief contrast and self-referential for the True > False Belief contrast) than to ToM maps. Fourth, no consistent brain behavioural correlations were present. Fifth, unlike prior work (Burnett and Blakemoore, 2009; Moessnang et al., 2017), we could not detect any increased functional connectivity between the temporo-parietal regions and the mPFC.

The main motivation for this study was to resolve prior discrepancy in fMRI studies of spontaneous ToM and to try to answer crucial open questions that require larger samples of participants (cf., Poldrack et al.,

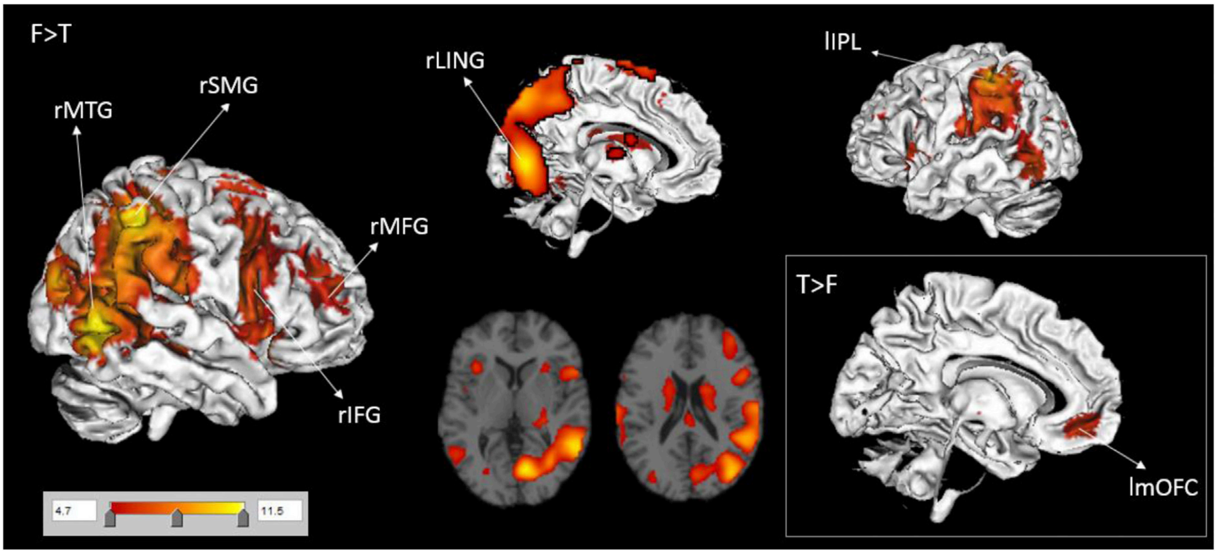

Fig. 1. Results of whole-brain analysis for the PxA analysis. Figures were visualised with MANGO software (http://ric.uthscsa.edu/mango/). The figure displays the main areas activated by the False Belief > True Belief contrast (F > T) and the True > False Belief contrast (T > F; bottom right panel) in the PxA analysis. The activated regions are coloured in red and yellow. SMG, supramarginal gyrus; MTG, middle temporal gyrus; LING, lingual gyrus; IFG, inferior frontal gyrus; MFG, middle frontal gyrus; IPL, inferior parietal lobule, mOFC, medial orbitofrontal cortex; r, right; l, left.

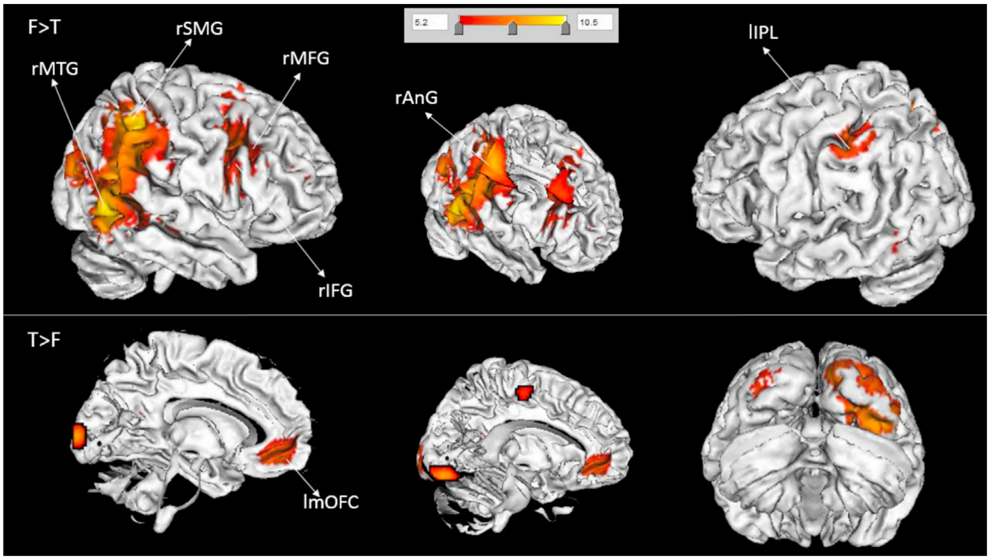

Fig. 2. Results of whole-brain analysis for the ToM Index analysis. The figure displays the main areas activated by the False Belief > True Belief contrast (top panel) and True > False Belief contrast (bottom panel) in the ToM Index analysis. The activated regions are coloured in red and yellow. SMG, supramarginal gyrus; MTG, middle temporal gyrus; IFG, inferior frontal gyrus; MFG, middle frontal gyrus; AnG, angular gyrus; IPL, inferior parietal lobule; mOFC, medial orbitofrontal cortex; r, right; l, left.

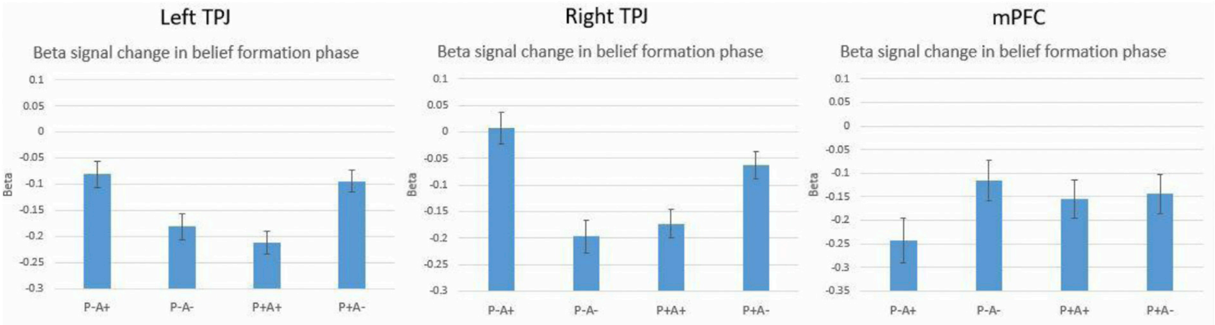

Fig. 3. ROI analysis. The graphs show the beta signal change in the belief formation phase for all the four different conditions (P-A+, P-A-, P+A+ and P + A-) in the left TPJ, right TPJ and mPFC ROIs respectively. The P-A+ and P + A-conditions are False Belief conditions and the P-A- and P+A+ are True Belief conditions. Error bars represent ±1 standard error.

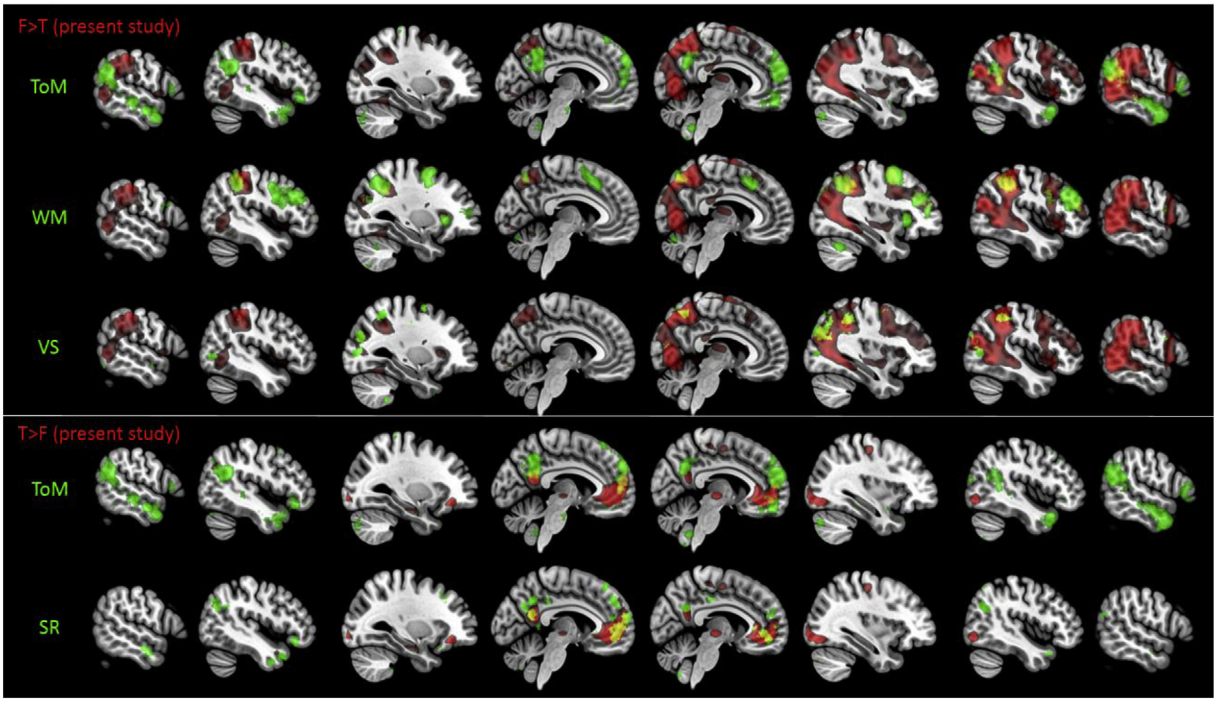

Fig. 4. Overlap with Neurosynth maps. Figures were visualised with MRIcroGL software (https://www.mccauslandcenter.sc.edu/mricrogl/). Red colour corresponds to our activation map for the PxA analysis in all the figures. First three rows from the top correspond to the False > True Belief (F > T) contrast, last two rows to the True > False (T > F) contrast. Green always corresponds to a map of Neurosynth. Yellow corresponds to the overlap between our map and Neurosynth map. From the top to the bottom: overlap with Theory of Mind (ToM) map, working memory map (WM), visuo-spatial map (VS), ToM and self-referential map (SR).

2017; Geuter et al., 2018). One contentious issue has been the involvement of the mPFC. Our results clearly indicate that the mPFC is not involved in false belief processing in a spontaneous ToM task, at least for the tracking of belief formation. This contrasts with substantial evidence that mPFC is consistently active in explicit ToM (Schurz et al., 2014; Van Overwalle, 2009) but is consistent with low reliability of this region in spontaneous ToM (Moessnang et al., 2016). Previous work in spontaneous ToM could not resolve the issue of whether or not the mPFC is activated because of low power and use of frequentist statistics. Our Bayesian analysis demonstrates that the mPFC is actually not recruited during spontaneous mentalizing. When considering the implications of these findings it is crucial to take into account that we measured brain activation in the belief formation phase while most explicit ToM studies measure activation when participants report the beliefs. It might well be that the mPFC is related to reporting beliefs, i.e. during later stages of cognitive processing, and that even in explicit ToM tasks the mPFC is not found to be active in the belief formation phase (e.g., Bardi et al., 2017). However, future studies will need to examine this issue in direct comparison.

By contrast, we previously reported TPJ activation in this sample across separate studies (Bardi et al., 2017; Nijhof et al., 2018), a finding that remained stable when pooling the populations together. Previous research on the role of the TPJ in spontaneous ToM has been contradictory (Bardi et al., 2017; Kovács et al., 2014; Schneider et al., 2014) and our results are in agreement with Kovács et al. (2014) but contrary to Schneider et al. (2014). The TPJ is a brain region consistently found to be active during explicit ToM and especially active in explicit false belief tasks (Döhnel et al., 2012; Saxe, 2010). Our finding provides confirming evidence for a role of TPJ in spontaneous ToM during the belief formation phase and the fact that the rTPJ was activated by inconsistencies between the belief of the participant and the belief of the agent.

Relatedly, a second central question that remained to be determined was whether this rTPJ activation was stronger in the right vs. the left side (in support of rTPJ dominance: Aichhorn et al., 2009; Döhnel et al., 2012; Liu et al., 2009; Saxe, 2010; in support of equal bilateral TPJ activation: Jenkins and Mitchell, 2010; Krall et al., 2015). Consistent with the main hypothesis, rTPJ relative to lTPJ activity was indeed statistically stronger although both regions were active during the False > True Belief condition. These findings are in agreement with previous reports of rTPJ dominance in explicit ToM (Aichhorn et al., 2009; Döhnel et al., 2012 Liu et al., 2009; Saxe, 2010) and spontaneous ToM (Bardi et al., 2017; Hudson et al., [preprint, bioRxiv]; Kovács et al., 2014; Nijhof et al., 2018), although not excluding an involvement of the lTPJ. Interestingly, this finding is also supported by the reliability analysis by Moessnang et al. (2016), who reported higher reliability of the right vs. the left temporo-parietal cortex during their abstract task of spontaneous mentalizing. Yet the precise underlying functional mechanisms of this lateralized activation pattern remain unknown.

Importantly, the TPJ was not the main source of brain activity and the main peaks of activation in the whole-brain analysis were located more dorsally and ventrally. The ventral peak corresponded to the middle temporal gyrus (MTG), a region implicated in ToM (Carrington and Bailey, 2009; Schurz et al., 2014) and involved in false belief processing (Rothmayr et al., 2011; Sommer et al., 2007; van Veluw and Chance, 2014). The dorsal peak corresponded to the supramarginal gyrus (SMG), or BA40, a region that is part of the inferior parietal lobule (IPL) (Igelström and Graziano, 2017) and is likewise activated in false belief tasks (see the meta-analysis by Schurz et al., 2013). However, the SMG is also involved in other functions, such as spatial working memory, spatial attention and visuospatial processing (Silk et al., 2010; Walter and Dassonville, 2008). Therefore, our findings, while confirming the activation of the TPJ in false belief reasoning, also extend this role to other parietal and temporal areas with peaks in the SMG and MTG, respectively. However, activation in frontal areas also emerged, namely, the right middle and inferior frontal gyri (MFG, IFG). These findings are in agreement with those reported in previous research showing MFG activity during both spontaneous (Naughtin et al., 2017) and explicit ToM (Rothmayr et al., 2011; Sommer et al., 2007). Like the SMG, the IFG is another area involved in attentional mechanisms, especially attention switching (Hedge et al., 2015), thus activating when a discrepancy in the observed situation catches our attention, such as when a belief of another person differs from our own.

To better integrate findings from the present analysis and to increase and facilitate comparability of our specific task and design with prior ToM work, we also compared the whole brain results to available maps collected by Neurosynth. This comparison revealed that activation obtained for the False > True Belief contrast corresponded most strongly to activation maps of visuospatial and working memory functions, especially in dorsal regions, and had only little resemblance with the typical activation pattern found in ToM tasks. The main dorsal peak of activation found in our analysis corresponded to the SMG, a region that, as said, is involved in spatial working memory and visuospatial processing (Silk et al., 2010; Walter and Dassonville, 2008). Since false belief processing requires individuals to keep in mind the belief of another person, working memory may be needed to correctly perform the task. Moreover, the task implemented in this study requires visuospatial processing, since participants have to track the location of a ball. When participants have a certain knowledge on the position of the ball that is different from that of the agent (false belief conditions), visuospatial areas might be recruited more because of the effort provided by knowing where the object actually is and simultaneously spontaneously imagining the ball in the location where the agent believes it to be. The dorsal peak of activation found in this study might therefore be involved in remembering and tracking the different assumed location of the ball, but still be important to correctly carry out the ToM task in the false belief conditions. As such, visuospatial functions may be recycled to monitor the beliefs of others (Corbetta et al., 2008) and so the dorsal activation could reflect real ToM activity and not be an artefact of the task.

The activation of regions involved in visuospatial processing, working memory and attention shifting, such as the SMG and the IFG, might suggest that spontaneous ToM actually is simply a domain-general spatial processing function, often referred to as “submentalizing”, that is activated when the observed situation presents a discrepancy between one’s own and another’s point of view, perspective or belief (see review by Heyes, 2014). This discrepancy would be processed in a spontaneous, fast and efficient way without the need to inhibit one’s own belief. Inhibition only comes into play when participants have to explicitly respond to a question during the task, thus requiring the involvement of the explicit ToM system, which includes the mPFC. However, more research into differences between different ToM accounts is needed.

Lastly, the behavioural analysis confirmed that the agent’s belief biased participants’ responses, which were faster when the agent believed the ball was present than when the agent believed the ball was absent. The correlation between behaviour in the outcome phase and brain activity in the belief formation phase, however, was not significant.

5. Limitations

Despite these intriguing findings, it is important to consider that the study did not include a functional explicit ToM localizer as comparison. Including a ToM localizer would be important to better elucidate how the pattern of activity in the temporo-parietal cortex found in the False > True Belief conditions overlapped with explicit ToM areas in a more direct way than with ToM maps in Neurovault and also to better define the ROIs, rather than using coordinates from a meta-analysis. Additionally, it is important to keep in mind that automated meta-analysis tools such as Neurosynth are not as accurate as manual meta-analyses, especially for specific and fine-grained cognitive domains. However, a validation study by Yarkoni et al. (2011) showed that Neurosynth is a valid and powerful approach for rapid large-scale high--quality meta-analyses involving broad domains (cf. also studies by Li et al., 2017 and Nummenmaa et al., 2018). Secondly, the present study used only one paradigm (the “Buzz Lightyear” task), which limits the ability to compare the different spontaneous ToM studies that have been conducted using different ToM tasks, such as the adapted version of the ‘Sally-Anne’ paradigm used by Schneider et al. (2014). However, by pooling data across studies, we nevertheless mitigated some of the power and sensitivity issues plaguing previous research. By doing so, the current study was able to resolve prior discrepancies in the spontaneous ToM literature, hence paving the way for future research to compare spontaneous and explicit ToM.

6. Conclusions

In conclusion, the present multi-study investigation sought to characterize the network involved in spontaneous ToM. The main analysis revealed a large cluster of activation related to spontaneous false belief processing in the right temporo-parietal cortex, encompassing the right TPJ and two main peaks in more dorsal and ventral locations, probably related to visuospatial and working memory functions that might be useful for carrying out the type of task used in this study or which might be inherent to ToM (see review by Corbetta et al., 2008). We, furthermore, confirmed the presence of an asymmetry in rTPJ versus lTPJ activation, but did not detect any mPFC activation or effects of functional connectivity. Finally, a comparison with available brain activation maps suggested overlap with a variety of higher-order cognitive functions. Thus, spontaneous ToM seems to be based on domain-general mechanisms, which provide a fast, efficient response to salient social stimuli, like discrepancies in belief or perspective, via low-level general spatial processing functions.

Funding

The study by Bardi et al. was funded by Research Foundation - Flanders (FWO) Pegasus Fellowship to Lara Bardi and by grant 331323-Mirroring and ToM, Marie Curie Fellowship (Marie Curie Intra-European fellowship for career development) to LB. The study by Nijhof et al. was supported by the Special Research Fund of Ghent University (project number BOF13/24J/083). The study by Hudson et al. was supported by a 2–4 year grant (01J05415) from the Special Research Fund (BOF) at Ghent University to SCM and MB.

Conflicts of interest

Declarations of interest: None.

Acknowledgements

Sara Boccadoro would like to thank the Erasmus + Traineeship program for the travel bursary.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuroimage.2019.116193.

References

Aichhorn, M., Perner, J., Weiss, B., Kronbichler, M., Staffen, W., Ladurner, G., 2009. Temporo-parietal junction activity in theory-of-mind tasks: falseness, beliefs, or attention. J. Cogn. Neurosci. 21 (6), 1179–1192. https://doi.org/10.1162/jocn.2009.21082.

Apperly, I.A., Samson, D., Chiavarino, C., Humphreys, G.W., 2004. Frontal and temporo-parietal lobe contributions to theory of mind: neuropsychological evidence from a false-belief task with reduced language and executive demands. J. Cogn. Neurosci. 16 (10), 1773–1784. https://doi.org/10.1162/0898929042947928.

Bardi, L., Desmet, C., Nijhof, A., Wiersema, J.R., Brass, M., 2017. Brain activation for spontaneous and explicit false belief tasks overlaps: new fMRI evidence on belief processing and violation of expectation. Soc. Cogn. Affect. Neurosci. 12 (3), 391–400. https://doi.org/10.1093/scan/nsw143.

Baron-Cohen, S., Leslie, A.M., Frith, U., 1985. Does the autistic child have a “theory of mind”? Cognition 21 (1), 37–46. https://doi.org/10.1016/0010-0277(85)90022-8.

Brett M, Anton J, Valabregue R, Poline J. Region of Interest Analysis Using an SPM Toolbox, Presented at the 8th International Conference on Functional Mapping of the Human Brain, June 26, 2002, Sendai, Japan, Available on CDROM in Neuroimage 16 (2).

Burnett, S., Blakemoore, S.J., 2009. Functional connectivity during a social emotion task in adolescents and adults. Eur. J. Neurosci. 29 (6), 1294–1301. https://doi.org/10.1111/j.1460-9568.2009.06674.x.

Carrington, S.J., Bailey, A.J., 2009. Are there theory of mind regions in the brain? A review of the neuroimaging literature. Hum. Brain Mapp. 30 (8), 2313–2335. https://doi.org/10.1002/hbm.20671.

Clements, W.A., Perner, J., 1994. Implicit understanding of belief. Cogn. Dev. 9 (4), 377–395. https://doi.org/10.1016/0885-2014(94)90012-4.

Corbetta, M., Patel, G., Shulman, G.L., 2008. The reorienting system of the human brain: from environment to theory of mind. Neuron 58 (3), 306–324. https://doi.org/10.1016/j.neuron.2008.04.017.

Decety, J., Lamm, C., 2007. The role of the right temporoparietal junction in social interaction: how low-level computational processes contribute to meta-cognition. The Neuroscientist 13 (6), 580–593. https://doi.org/10.1177/1073858407304654.

Deschrijver, E., Bardi, L., Wiersema, J.R., Brass, M., 2016. Behavioral measures of implicit theory of mind in adults with high functioning autism. Cogn. Neurosci. 7 (1–4), 192–202. https://doi.org/10.1080/17588928.2015.1085375.

Döhnel, K., Schuwerk, T., Meinhardt, J., Sodian, B., Hajak, G., Sommer, M., 2012. Functional activity of the right temporo-parietal junction and of the medial prefrontal cortex associated with true and false belief reasoning. Neuroimage 60 (3), 1652–1661. https://doi.org/10.1016/j.neuroimage.2012.01.073.

Durfour, N., Redcay, E., Young, L., Mavros, P.L., Moran, J.M., Triantafyllou, C., Gabrieli, J.D.E., Saxe, R., 2013. Similar brain activation during false belief tasks in a large sample of adults with and without autism. PLoS One (9) e75468. https://doi.org/10.1371/journal.pone.0075468.

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

Gallagher, H.L., Happe, F., Brunswick, N., Fletcher, P.C., Frith, U., Frith, C.D., 2000. Reading the mind in cartoons and stories: an fMRI study of 'theory of mind' in verbal and nonverbal tasks. Neuropsychologia 38 (1), 11–21.

Geuter, S., Qi, G., Welsh, R.C., Wager, T.D., Lindquist, M.A., 2018. Effect Size and Power in fMRI Group Analysis bioRxiv 295048. https://doi.org/10.1101/295048.

Gorgolewski, K.J., Varoquaux, G., Rivera, G., Schwarz, Y., Ghosh, S.S., Maumet, C., et al., 2015. NeuroVault.org: a web-based repository for collecting and sharing unthresholded statistical maps of the human brain. Front. Neuroinf. 9, 8. https://doi.org/10.3389/fninf.2015.00008.

Hedge, C., Stothart, G., Todd Jones, J., Rojas Frias, P., Magee, K.L., Brooks, J.C., 2015. A frontal attention mechanism in the visual mismatch negativity. Behav. Brain Res. 293, 173–181. https://doi.org/10.1016/j.bbr.2015.07.022.

Heyes, C., 2014. Submentalizing: I'm not really reading your mind. Psychol. Sci. 9, 121–143. https://doi.org/10.1177/1745691613518076.

Hudson, A.R., Van Hamme, C., Maeyens, L., Brass, M., Mueller, S.C.. Spontaneous mentalizing after early interpersonal trauma: evidence for hypoactivation of the temporoparietal junction [Preprint], bioRxiv ID ID 487363. https://doi.org/10.1101/487363.

Igelström, K.M., Graziano, M.S.A., 2017. The inferior parietal lobule and temporoparietal junction: a network perspective. Neuropsychologia 105, 70–83. https://doi.org/10.1016/j.neuropsychologia.2017.01.001.

Jenkins, A.C., Mitchell, J.P., 2010. Mentalizing under uncertainty: dissociated neural responses to ambiguous and unambiguous mental state inferences. Cerebr. Cortex 20 (2), 404–410. https://doi.org/10.1093/cercor/bhp109.

Kovács, A.M., Kühn, S., Gergely, G., Csibra, G., Brass, M., 2014. Are all beliefs equal? Implicit belief attributions recruiting core brain regions of theory of mind. PLoS One 9 (9). https://doi.org/10.1371/journal.pone.0106558 e106558.

Kovács, A.M., Téglás, E., Endress, A.D., 2010. The social sense: susceptibility to others' beliefs in human infants and adults. Science 330 (6012), 1830–1834. https://doi.org/10.1126/science.1190792.

Krall, S.C., Rottschy, C., Oberwelland, E., Bzdok, D., Fox, P.T., Eickhoff, S.B., et al., 2015. The role of the right temporoparietal junction in attention and social interaction as revealed by ALE meta-analysis. Brain Struct. Funct. 220 (2), 587–604. https://doi.org/10.1007/s00429-014-0803-z.

Li, R., Smith, D.V., Clithero, J.A., Venkatraman, V., Carter, R.M., Huetterl, S.A., 2017. Reason's enemy is not emotion: engagement of cognitive control networks explains biases in gain/loss framing. J. Neurosci. 37 (13), 3588–3598. https://doi.org/10.1523/JNEUROSCI.3486-16.2017.

Liu, D., Meltzoff, A.N., Wellman, H.M., 2009. Neural correlates of belief- and desire-reasoning. Child Dev. 80 (4), 1163–1171. https://doi.org/10.1111/j.1467-8624.2009.01323.x. Epub 2009/07/28.

Maxwell, S.E., Lau, M.Y., Howard, G.S., 2015. Is psychology suffering from a replication crisis? What does “failure to replicate” really mean?, 70 (6), 487–498. https://doi.org/10.1037/a0039400.

McLaren, D.G., Ries, M.L., Xu, G., Johnson, S.C., 2012. A generalized form of context-dependent psychophysiological interactions (gPPI): a comparison to standard approaches. Neuroimage 61 (4), 1277–1286. https://doi.org/10.1016/j.neuroimage.2012.03.068.

Molenberghs, P., Johnson, H., Henry, J.D., Mattingley, J.B., 2016. Understanding the minds of others: a neuroimaging meta-analysis. Neurosci. Biobehav. Rev. 65, 276–291. https://doi.org/10.1016/j.neubiorev.2016.03.020.

Moessnang, C., Schäfer, A., Bilek, E., Roux, P., Otto, K., Baumeister, S., Hohmann, S., Poustka, L., Brandeis, D., Banaschewski, T., Meyer-Lindenberg, A., Tost, H., 2016. Specificity, reliability and sensitivity of social brain responses during spontaneous mentalizing. Soc. Cogn. Affect. Neurosci. 11 (11), 1687–1697.

Moessnang, C., Otto, K., Bilek, E., Schäfer, A., Baumeister, S., Hohmann, S., Poustka, L., Brandeis, D., Banaschewski, T., Tost, H., Meyer-Lindenberg, A., 2017. Differential responses of the dorsomedial prefrontal cortex and right posterior superior temporal sulcus to spontaneous mentalizing. Hum. Brain Mapp. 38 (8), 3791–3803.

Naughtin, C.K., Horne, K., Schneider, D., Venini, D., York, A., Dux, P.E., 2017. Do implicit and explicit belief processing share neural substrates? Hum. Brain Mapp. 38 (9), 4760–4772. https://doi.org/10.1002/hbm.23700.

Nijhof, A.D., Bardi, L., Brass, M., Wiersema, J.R., 2018. Brain activity for spontaneous and explicit mentalizing in adults with autism spectrum disorder: an fMRI study. Neuroimage Clin. 18, 475–484. https://doi.org/10.1016/j.nicl.2018.02.016.

Nummenmaa, L., Hari, R., Hietanen, J.K., Glerean, E., 2018. Maps of subjective feelings. Proc. Natl. Acad. Sci. U.S.A. 115 (37), 9198–9203. https://doi.org/10.1073/pnas.1807390115.

Onishi, K.H., Baillargeon, R., 2005. Do 15-month-old infants understand false beliefs? Science 308 (5719), 255–258. https://doi.org/10.1126/science.1107621.

Poldrack, R.A., Baker, C.I., Durnez, J., Gorgolewski, K.J., Matthews, P.M., Munafo, M.R., Nichols, T.E., Poline, J.B., Vul, E., Yarkonio, T., 2017. Scanning the Horizon: towards Transparent and Reproducible Neuroimaging Research, vol. 18, pp. 115–126. https://doi.org/10.1038/nrn.2016.167 (2).

Rothmayr, C., Sodian, B., Hajak, G., Döhnel, K., Meinhardt, J., Sommer, M., 2011. Common and distinct neural networks for false-belief reasoning and inhibitory control. Neuroimage 56 (3), 1705–1713. https://doi.org/10.1016/j.neuroimage.2010.12.052.

Rouder, J.N., Speckman, P.L., Sun, D., Morey, R.D., Iverson, G., 2009. Bayesian t tests for accepting and rejecting the null hypothesis. Psychon. Bull. Rev. 16 (2), 225–237.

Ruby, P., Decety, J., 2003. What you believe versus what you think they believe: a neuroimaging study of conceptual perspective-taking. Eur. J. Neurosci. 17 (11), 2475–2480.

Samson, D., Apperly, I.A., Chiavarino, C., Humphreys, G.W., 2004. Left temporoparietal junction is necessary for representing someone else's belief. Nat. Neurosci. 7 (5), 499–500. https://doi.org/10.1038/nn1223.

Saxe, R., 2010. The right temporo-parietal junction: a specific brain region for thinking about thoughts. In: Leslie, A., German, T. (Eds.), Handbook of Theory of Mind.

Saxe, R., Kanwisher, N., 2003. People thinking about thinking people. The role of the temporo-parietal junction in “theory of mind”. Neuroimage 19 (4), 1835–1842.

Schneider, D., Slaughter, V.P., Becker, S.I., Dux, P.E., 2014. Implicit false-belief processing in the human brain. Neuroimage 101, 268–275. https://doi.org/10.1016/j.neuroimage.2014.07.014.

Schneider, D., Slaughter, V.P., Dux, P.E., 2017. Current evidence for automatic Theory of Mind processing in adults. Cognition 162, 27–31. https://doi.org/10.1016/j.cognition.2017.01.018.

Schurz, M., Aichhorn, M., Martin, A., Perner, J., 2013. Common brain areas engaged in false belief reasoning and visual perspective taking: a meta-analysis of functional brain imaging studies. Front. Hum. Neurosci. 7, 712. https://doi.org/10.3389/fnhum.2013.00712.

Schurz, M., Radua, J., Aichhorn, M., Richlan, F., Perner, J., 2014. Fractionating theory of mind: a meta-analysis of functional brain imaging studies. Neurosci. Biobehav. Rev. 42, 9–34. https://doi.org/10.1016/j.neubiorev.2014.01.009.

Silk, T.J., Bellgrove, M.A., Wrafter, P., Mattingley, J.B., Cunnington, R., 2010. Spatial working memory and spatial attention rely on common neural processes in the intraparietal sulcus. Neuroimage 53 (2), 718–724. https://doi.org/10.1016/j.neuroimage.2010.06.068.

Sommer, M., Döhnel, K., Sodian, B., Meinhardt, J., Thoermer, C., Hajak, G., 2007. Neural correlates of true and false belief reasoning. Neuroimage 35 (3), 1378–1384. https://doi.org/10.1016/j.neuroimage.2007.01.042.

Southgate, V., Senju, A., Csibra, G., 2007. Action anticipation through attribution of false belief by 2-year-olds. Psychol. Sci. 18 (7), 587–592. https://doi.org/10.1111/j.1467-9280.2007.01944.x.

Van Overwalle, F., 2009. Social cognition and the brain: a meta-analysis. Hum. Brain Mapp. 30 (3), 829–858. https://doi.org/10.1002/hbm.20547.

van Veluw, S.J., Chance, S.A., 2014. Differentiating between self and others: an ALE meta-analysis of fMRI studies of self-recognition and theory of mind. Brain Imag. Behav. 8 (1), 24–38. https://doi.org/10.1007/s11682-013-9266-8.

Walter, E., Dassonville, P., 2008. Visuospatial contextual processing in the parietal cortex: an fMRI investigation of the induced Roelofs effect. Neuroimage 42 (4), 1686–1697. https://doi.org/10.1016/j.neuroimage.2008.06.016.

Wellman, H.M., Cross, D., Watson, J., 2001. Meta-analysis of theory-of-mind development: the truth about false belief. Child Dev. 72 (3), 655–684.

Wimmer, H., Perner, J., 1983. Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. Cognition 13 (1), 103–128. https://doi.org/10.1016/0010-0277(83)90004-5.

Yarkoni, T., Poldrack, R.A., Nichols, T.E., Van Essen, D.C., Wager, T.D., 2011. Large-scale automated synthesis of human functional neuroimaging data. Nat. Methods 8 (8), 665–670 https://dx.doi.org/10.1038%2Fnmeth.1635.