



## Review

## Fractionating theory of mind: A meta-analysis of functional brain imaging studies



Matthias Schurz <sup>a,b,\*</sup>, Joaquim Radua <sup>c,d</sup>, Markus Aichhorn <sup>a,b</sup>,  
Fabio Richlan <sup>a,b</sup>, Josef Perner <sup>a,b</sup>

<sup>a</sup> Department of Psychology, University of Salzburg, Hellbrunnerstr. 34, 5020 Salzburg, Austria

<sup>b</sup> Center for Neurocognitive Research, University of Salzburg, Hellbrunnerstr. 34, 5020 Salzburg, Austria

<sup>c</sup> Institute of Psychiatry, King's College London, De Crespigny Park, London SE5 8AF, United Kingdom

<sup>d</sup> FIDMAG Germanes Hospitalaries Research Unit, CIBERSAM, Sant Boi de Llobregat, Barcelona, Spain

## ARTICLE INFO

## Article history:

Received 9 December 2013

Received in revised form 17 January 2014

Accepted 20 January 2014

## Keywords:

Theory of mind  
Mentalizing  
Meta-analysis  
Task groups  
Medial prefrontal cortex  
Precuneus  
Temporo-parietal junction  
Anterior temporal lobe  
Inferior frontal gyrus

## ABSTRACT

We meta-analyzed imaging studies on theory of mind and formed individual task groups based on stimuli and instructions. Overlap in brain activation between all task groups was found in the mPFC and in the bilateral posterior TPJ. This supports the idea of a core network for theory of mind that is activated whenever we are reasoning about mental states, irrespective of the task- and stimulus-formats (Mar, 2011). In addition, we found a number of task-related activation differences surrounding this core-network. ROI based analyses show that areas in the TPJ, the mPFC, the precuneus, the temporal lobes and the inferior frontal gyri have distinct profiles of task-related activation. Functional accounts of these areas are reviewed and discussed with respect to our findings.

© 2014 The Authors. Published by Elsevier Ltd. All rights reserved.

## Contents

1. Introduction .....	10
2. Methods.....	11
2.1. Literature search and study selection .....	11
2.2. Meta-analytic methods .....	11
2.2.1. Effect-size signed differential mapping.....	11
2.2.2. Complementary analyses .....	11
2.2.3. Overlap analyses .....	12
2.2.4. ROI analyses.....	12
2.2.5. Task group comparisons .....	12
2.3. Definition of TPJ and mPFC .....	12
3. Results .....	13
3.1. Individual meta-analyses .....	13
3.1.1. False belief vs. photo .....	13
3.1.2. Trait Judgments .....	13
3.1.3. Strategic games .....	15
3.1.4. Social animations.....	16
3.1.5. Mind in the eyes .....	19
3.1.6. Rational actions .....	19
3.1.7. Summarizing convergent activations across task groups .....	19

\* Corresponding author at: Center for Neurocognitive Research, University of Salzburg, Hellbrunnerstr. 34, 5020 Salzburg, Austria. Tel.: +43 662 8044 5142; fax: +43 662 8044 5126.

E-mail address: [matthias.schurz@sbg.ac.at](mailto:matthias.schurz@sbg.ac.at) (M. Schurz).

3.2.	Region of interest (ROI) analyses .....	21
3.2.1.	Temporo-parietal ROIs .....	23
3.2.2.	Cortical midline ROIs .....	23
3.2.3.	Fronto-temporal ROIs .....	23
4.	Discussion .....	24
4.1.	The role of conceptual variance for meta-analysing theory of mind .....	24
4.2.	Contribution of the present meta-analysis .....	25
4.3.	Core network for theory of mind .....	26
4.4.	Areas for theory of mind .....	26
4.4.1.	Temporo-parietal junction .....	26
4.4.2.	Medial prefrontal cortex .....	27
4.4.3.	Precuneus .....	29
4.4.4.	Anterior temporal lobes .....	29
4.4.5.	Inferior frontal gyrus .....	29
4.5.	Limitations .....	30
5.	Conclusion .....	30
	Appendix A. Supplementary data .....	30
	References .....	30

## 1. Introduction

A central component of human social cognition is the ability to attribute mental states to ourselves and others. This ability is often referred to as ‘mentalizing’, ‘mindreading’ or ‘theory of mind’. For nearly two decades, functional imaging has been used to study how this ability is implemented in the brain. To date, hundreds of empirical studies on this topic can be found in the literature. In the present study, we meta-analyze these imaging findings by forming different task groups based on the stimuli and instructions used.

Traditionally, the literature has been summarized in narrative reviews (e.g., Saxe, 2006; Frith and Frith, 2006). Some reviews proposed that all sorts of theory of mind tasks consistently engage a particular brain network (e.g., Amodio and Frith, 2006; Frith and Frith, 2006; Mitchell, 2009). This network includes the medial pre-frontal cortex (mPFC), parts of precuneus and posterior cingulate cortex, the temporo-parietal junction (TPJ) and the posterior superior temporal sulcus (pSTS) bilaterally. Less frequently, the bilateral anterior temporal lobes and areas in the amygdala have also been associated with this network (Frith and Frith, 2006; Mar, 2011). The common feature of all tasks activating this network is that they require thinking about the mental states of other persons (Frith and Frith, 2006). For example, Gallagher and Frith (2003, p. 78) noted that functional imaging studies on mentalizing have ‘... demonstrated remarkably consistent results despite using multimodal and diverse cognitive paradigms, such as verbal and non-verbal or on- and off-line tasks. Similarly, Mitchell (2009, p. 1310) noted: ‘... studies have established a small and highly reliable network of regions that is preferentially engaged when perceivers mentalize about the mind of others.’ The author added ‘...researchers observed this pattern of activation regardless of whether mental inferences are prompted by stories ..., cartoons ..., in the context of competitive and economic games ..., or by task instructions to think about a specific person’s mind ...’.

At the same time, many researchers say that particular brain areas subserve particular sub-processes of theory of mind (e.g., Frith and Frith, 1999; Saxe et al., 2004a, 2004b; Gallagher and Frith, 2003; Amodio and Frith, 2006; Gobbini et al., 2007; Perner and Leekam, 2008). Moreover, it has been proposed that some sub-processes of theory of mind are preferentially engaged in some tasks and not in others (e.g., Aichhorn et al., 2006; Bahnemann et al., 2010; Gobbini et al., 2007; Perner and Leekam, 2008). For example, Gobbini et al. (2007) observed that theory of mind tasks that focus on false beliefs tend to activate the TPJ more dorsally than tasks without false beliefs (e.g., social animations and point-light movement perception), which activate more ventrally in the pSTS.

In the recent years, quantitative meta-analyses have become a tool for validating and extending the findings of narrative imaging reviews. They offer a statistical procedure to objectively summarize imaging findings in the form of brain-maps of convergent activation (e.g., Eickhoff et al., 2009; Radua et al., 2012). Several groups (e.g., Bzdok et al., 2012; Decety and Lamm, 2007; Spreng et al., 2009) performed ‘pooled’ meta-analyses on theory of mind, i.e., putting together all published imaging studies that used a term like ‘theory of mind’ in their keywords. Findings of these meta-analyses show the same large-scale network of areas that has been described in narrative reviews. A methodological strength of pooled meta-analyses is that they have high statistical power to detect findings, as they are based on a large sample of all available studies. Large variability, however, is also a drawback of these approaches, as they put together studies using different stimulus material and control conditions, as well as studies with variable methodological quality and rigor. In a conceptually diverse field such as theory of mind research, this will produce a lot of noise. For example, Mar (2011) noted that: ‘Differences in approach to studying ToM may explain some of the current disagreement regarding which areas are key for mentalizing’ (p. 117). Up to now, only one review has looked at task-related differences in activation for theory of mind with a quantitative method (Van Overwalle, 2009). The author performed a region of interest based meta-analysis on social cognition, including the domain of theory of mind. A limitation of van Overwalle’s (2009) approach is that findings of his method are bound to the scope of the a priori defined regions of interest. For example, the TPJ was analyzed with one single region of interest, spanning from pSTS (at  $z=3$ ) up to the dorsal TPJ (located in the inferior parietal lobe, at  $z=39$ ). The author observed convergent activation in this region for a wide variety of theory of mind tasks. For example, consistent activation was found in the TPJ for false belief tasks as well as for social animations. This observation stands in contrast to what was proposed by Gobbini et al. (2007) and other researchers (e.g., Aichhorn et al., 2006; Bahnemann et al., 2010; Perner and Leekam, 2008).

In the present study, we performed a review of theory of mind imaging studies, and sorted them into task groups that had comparable stimulus-material, instructions and control conditions. This reduces the problem of pooling across studies which differ on conceptual and methodological aspects, and offers a more-fine grained picture of the evidence. In addition, we performed a whole-brain meta-analysis on a voxel-by-voxel basis, which allowed us to look at local differences in more detail than earlier works (e.g., Van Overwalle, 2009). Our meta-analysis aims at providing a systematic and quantitative evaluation of two previous claims. First, we

wanted to evaluate to what extent different theory of mind tasks consistently activate a common set of brain areas (e.g., Amodio and Frith, 2006; Frith and Frith, 2006; Mitchell, 2009). Second, we wanted to evaluate if some sub-areas of the theory of mind brain network are preferentially engaged by some tasks and not by others (Aichhorn et al., 2006; Bahne et al., 2010; Gobbini et al., 2007; Perner and Leekam, 2008).

## 2. Methods

### 2.1. Literature search and study selection

The literature reviewed in our meta-analysis was retrieved by several strategies. First, we adopted the literature sample of a recent large-scale meta-analysis (Mar, 2011). The author categorized studies into story-based versus non-story based tasks, and looked at meta-analytic overlaps between them. In contrast, the present study seeks to meta-analyze brain activation for individual groups formed on the basis of a certain type of task, i.e., studies that have presented comparable sorts of stimuli and instructions. Mar's meta-analysis contained literature on theory of mind published between 1995 and 2010 with the following strategy: Mar (2011) performed a key-word search in the databases PubMed, Science Citation Index, and PsycInfo. The criteria were the following: studies were selected if (i) they included one of the key-words 'neuroimaging' or 'fMRI' or 'PET' and (ii) one of the key-words 'theory-of-mind' or 'mentalizing' or 'mindreading'. We updated Mar's (2011) literature sample by replicating the author's search for the literature published between 2010 and 2013 (up to January 2013). Based on the literature sample retrieved in these two steps, we attempted to identify task groups representing frequently used experimental paradigms to study theory of mind. A task group had to contain studies which used similar stimuli and presentation-modes, similar task-instructions, and results from comparable contrasts. As we are not aware of exact criteria for the minimum sample size in imaging meta-analysis, we ensured that our task groups had samples within the typical range found in the literature. We refer to a recent summary article of 94 published imaging meta-analyses (David et al., 2013). This paper reports that the interquartile-range of the number of participants included in imaging meta-analysis lies between 123 and 355, and the IQR for the number of original studies included lies between 8 and 20. All our task groups had sample-sizes well within that range. After a preliminary list of task groups was defined, we searched through multiple existing literature reviews on theory of mind and related topics to find additional studies falling into these groups (Bzdok et al., 2012; Denny et al., 2012; Murray et al., 2012; Perner and Leekam, 2008; Spreng et al., 2009; Van Overwalle, 2009; Van Overwalle and Baetens, 2009).

We then applied a number of methodological selection-criteria to the body of literature identified by our search (see e.g., Radua et al., 2012). Studies were only selected if they had performed a whole brain analysis and reported activation coordinates in standard space (MNI or Talairach). Data from clinical samples were removed. We ensured that the same threshold throughout the whole brain was used within each included study, in order to avoid biases toward liberally thresholded brain regions. Note that this does not mean that different studies should employ the same threshold. We also ensured that all activation maps entering our meta-analysis refer to a basic process of reasoning about/processing mental states. Therefore, we only selected studies which contrasted a task clearly referring to mental state processing versus a task which clearly did not. This excluded several study types, for example studies which compared inferences about others' versus own mental states or which did not engage mental states in any condition. If a study reported more than one contrast, we

selected the one best corresponding to contrasts reported in other studies of a task group. In total, we meta-analyzed 757 activation foci reported from 73 studies that tested 1241 participants. Our meta-analysis contained six different task groups: (i) false belief, vs. photo, including 112 activation foci from 15 studies with 259 participants, (ii) trait judgment tasks, 111 foci, 15 studies, 253 participants, (iii) strategic games, 68 foci, 9 studies, 162 participants, (iv) social animations, 203 foci, 14 studies, 224 participants, (v) mind in the eyes tasks, 146 foci, 10 studies, 185 participants and (vi) rational actions, 117 foci, 10 studies, 158 participants.

### 2.2. Meta-analytic methods

#### 2.2.1. Effect-size signed differential mapping

We used Effect-Size Signed Differential Mapping (ES-SDM) software, version 2.31 for meta-analysis (Radua et al., 2012; [www.sdmproject.com](http://www.sdmproject.com)). SDM is based and improves upon the positive features from existing peak probability methods for meta-analysis, such as Activation Likelihood Estimation (ALE, Eickhoff et al., 2009) or Multilevel Kernel Density Analysis (MKDA, Wager et al., 2007). ES-SDM uses standard effect size and variance-based meta-analytic calculations. Based on the reported *t*-values and the sample size of a study, ES-SDM creates a map of effect-sizes (Hedge's *g* values) and a map of variance, with the latter being derived from the distribution of effect-sizes and the sample size of the study. Effect-sizes are calculated for those voxels containing a peak that is reported in the results table of an original study. For the remaining voxels, an effect-size is estimated depending on the distance to close peaks (<20 mm) by means of an unnormalized Gaussian kernel. In the present analysis, we used the recommended Gaussian kernel with a FWHM of 20 mm. A validation study which compared the results of coordinate based ES-SDM meta-analysis to the results of a standard voxel-wise GLM analysis of the same original data (Radua et al., 2012) found that this FWHM provided an optimal balance between sensitivity and specificity. For statistical-analysis, all foci were transformed to Talairach space which is the native space of the software, by using the matrix transformations proposed by Lancaster et al. (2007).

We calculated a mean analysis for each task group. Calculation of the meta-analytic mean map is implemented by a random-effects model in which each study is weighted by the inverse of the sum of its variance plus an estimate of between-study heterogeneity. The latter is obtained by the DerSimonian–Laird method (DerSimonian and Laird, 1986). This approach enables studies with larger sample size or lower variability to contribute more, and that effects are assumed to randomly vary between samples. Statistical significance of the mean map was assessed by a permutation test that randomizes the location of the voxels within the standard SDM gray matter template. Thus, the null hypothesis of these permutation tests is that effect-sizes are randomly distributed throughout the brain. 100 random maps were generated with the same number of input foci as included in the to-be-tested map. Finally, the meta-analytic maps were thresholded using a voxel-level (height) threshold of  $p < 0.005$  (uncorrected) and a cluster-level (extent) threshold of 10 voxels. This uncorrected threshold was found to optimally balance sensitivity and specificity, and to be an approximate equivalent to a corrected threshold of  $p < 0.05$  in original neuroimaging studies (Radua et al., 2012). For convenience, we report all activations in MNI-space.

#### 2.2.2. Complementary analyses

We used the heterogeneity analysis in ES-SDM to check whether the brain areas found by our mean analysis showed strong variability across studies. ES-SDM calculates a *Q*-statistic based on the between study variance in effect-size estimates for a given area. Heterogeneity can then be tested for significance by determining if

the observed between-study variance for a given area is larger than that resulting from sampling error alone. For convenience, heterogeneity values are converted into standard  $z$  values in ES-SDM.

Systematic whole-brain voxel-based jackknife sensitivity analysis was conducted. This consists of repeating the main statistical analysis for each study in the meta-analysis once, and discarding a different study each time. This means, we removed one study at a time, performed an analysis on the remaining studies, replaced the study and repeated this process. The rationale of this test is that if a previously significant brain region remains significant in all or most of the combinations of studies, it can be concluded that this finding is highly replicable.

### 2.2.3. Overlap analyses

We used three methods to determine brain regions that were activated by all task groups. First, we used conjunction analysis (Nichols et al., 2005), which calculates a simple overlap between regions that were found statistically significant in the individual meta-analyses. This was done with the 'image calculator' utility in SPM8 ([www.fil.ion.ucl.ac.uk](http://www.fil.ion.ucl.ac.uk)). This procedure shows findings which are individually significant in all six meta-analyses at our default threshold of  $p < .005$  uncorrected,  $z > 1$ , and minimum cluster extent 10 voxels.

However, the conjunction approach may be biased toward a false impression of selectivity. This is because the number of studies in each of the task-specific meta-analyses is much lower than the number of studies in the pooled analysis, thus reducing the power of the individual analyses. To address this problem, we carried out a permutation based overlap analysis. Here, we determined the overlap between two halves of the pooled meta analysis sample, each consisting of the studies from three different task groups. This procedure was repeated ten times to create every permutation for assigning six task groups to two sample-halves. Carrying out conjunction analyses between the sample-halves yielded ten maps of overlap, which were then again merged by conjunction analysis, giving a final map of overlap between task groups.

The third way to determine overlaps between task groups was a multimodal meta-analysis as described by Radua et al. (2013). Whereas the conjunction analysis determines the overlap between binary thresholded maps, multimodal meta-analysis determines overlap based on the actual  $p$ -values from the original meta-analytic maps. Thereby, the method takes into account that  $p$ -values from the unimodal meta-analyses may be estimated with some degree of error. Significance is determined by an adjusted  $U$  statistic based on the convenient null-hypothesis that not a single one of the meta-analyses shows activation in an area (for more details, see Radua et al., 2013). The multimodal meta-analysis was initially created for the case of two meta-analyses (e.g., Radua et al., 2012; Cooper et al., 2013) and it was noted that the method should be conducted with caution in the case of more than two meta-analyses, as it may become too liberal. Conjunction analysis, on the other hand, becomes increasingly conservative in the case of more than two individual meta-analyses.

### 2.2.4. ROI analyses

We focused on ROI analyses to characterize the task-specific patterns of activation underlying the main findings of our pooled-analysis. Therefore, we placed ROIs at the locations of peak voxels from our pooled analysis (highlighted in bold in Table 3). For large clusters (temporo-parietal clusters, medial prefrontal cluster), we additionally placed ROIs at local subpeaks, according to the following criteria: (i) Centers of all ROIs had a minimum distance of 20 mm, which corresponds to the FWHM of meta-analytic smoothing. (ii) ROIs were placed in different functional-anatomical areas or subdivisions of areas (e.g., the TPJ), as described by brain atlases. For meta-analytic clusters that were found both in the left and

right hemisphere, we started with ROI selection on the right side, and then chose closest possible peak or sub-peak on the homologous left coordinate. For the TPJ and the mPFC we relied on recent connectivity-based parcellation atlases (Mars et al., 2011, 2012; Sallet et al., 2013) for labeling our results and for selection of ROIs. These atlases subdivide the TPJ and the mPFC based on patterns of structural and functional connectivity (for details see Section 2.3).

For locating other ROIs in our analysis, such as precuneus and fronto-temporal ROIs, we relied on classical macroanatomic atlas boundaries (Talairach Daemon). The only exceptions to this procedure were our ROIs in the anterior temporal lobes. We selected the anterior temporal ROIs because they were located close to the temporal poles, for which we could not find proper local subpeaks. For one finding from our pooled meta-analysis, the left fusiform gyrus, we did not analyze ROI data because of its closeness to another ventral temporal ROI (left middle temporal ROI). For all ROIs, we show boxplots that illustrate the distribution of effect-size estimates within each task group. Additionally, we report the results of the meta-analytic mean analysis for our ROIs, i.e., results from the analysis that was described earlier for calculating meta-analytic mean maps. Significance is reported with our default threshold of  $p < .005$ ,  $z > 1$ . For interest, we also report statistical trends found at a more liberal threshold of  $p < .05$ ,  $z > 1$ .

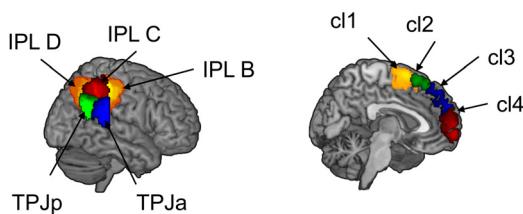
### 2.2.5. Task group comparisons

To test for differences between the meta-analyses of individual task groups, we carried out linear-models in ES-SDM. Linear models calculate the difference between effect-size estimates from two meta-analyses while taking into account differences in sample size as well as within- and between study variability. This method ensures that differences between groups are not driven by differences in sample size or heterogeneity between the meta-analyses. We performed pairwise comparisons for all 15 combinations of the six task groups in our main analysis. We focus our results on the ROIs which were described in the last section. To account for the multiple comparison problem, we corrected our default threshold of  $p < .005$ ,  $z > 1$  with the Bonferroni method, giving  $p < .00033$ ,  $z > 1$ . For interest, we also report group differences that are significant only at the uncorrected default threshold  $p < .005$ ,  $z > 1$ .

### 2.3. Definition of TPJ and mPFC

Two of the most commonly mentioned brain structures in theory of mind research are the temporo-parietal junction (TPJ) and the medial prefrontal cortex (mPFC). Both of these labels are only poorly described by classical macroanatomic boundaries and have been used somewhat inconsistently. To characterize and discuss the results from our meta-analysis for the TPJ and the mPFC we rely on recent atlases that map these areas and their subregions based on structural connectivity profiles obtained from diffusion weighted imaging. These atlases are available in fslview (<http://fsl.fmrib.ox.ac.uk/fsl/fslview/>). Fig. 1 illustrates the connectivity-parcellation clusters of the TPJ and the mPFC. Mars et al. (2012) found that a more anterior connectivity cluster of the TPJ, labeled as TPJa, is interconnected with the inferior frontal gyrus, anterior insula and the SMA. A more posterior connectivity cluster, the TPJp, is interconnected with more dorsal-posterior areas in the IPL, the precuneus, the ventral mPFC, and posterior middle temporal areas. Yet another profile was found for the connectivity cluster in the IPL (dorsal to the TPJa and TPJp), which is, for example, more strongly connected with lateral prefrontal areas. The IPL connectivity cluster can again be subdivided into a number of subclusters along a posterior-to-anterior gradient (Mars et al., 2011). Most relevant for our work are connectivity cluster IPL B, IPL C and IPL D, which lie directly above the TPJ connectivity clusters. Sallet et al. (2013) used connectivity based parcellation

## Connectivity based parcellations of TPJ and mPFC



**Fig. 1.** Illustration of structural connectivity-based mappings of TPJ and mPFC. All maps were thresholded at 25% of participants, so only voxels that belong to any given mask in 25–100% of the participants are labeled. The TPJ was divided in posterior (TPJp), anterior (TPJa), and dorsal (IPL) clusters by Mars et al. (2012). The IPL connectivity cluster can again be subdivided into a number of subclusters (Mars et al., 2011), most relevant for our work are connectivity cluster IPL B, IPL C and IPL D, which lie directly above the TPJ connectivity clusters. For the mPFC, we show connectivity clusters 1–4 derived from a parcellation of the frontal cortex by Sallet et al. (2013).

to divide the medial frontal cortex into several connectivity clusters. Most relevant for our results are medial frontal connectivity clusters 2–4. Connectivity-cluster 2 corresponds to the pre-SMA area and shows strong connectivity to widespread areas in more anterior medial prefrontal areas and in lateral prefrontal areas. Connectivity-clusters 3 and 4 lie in the mPFC, and cluster 4 lies ventral to cluster 3. Both connectivity clusters 3 and 4 show strong a linkage to areas throughout the medial prefrontal cortex, the posterior cingulate gyrus, bilateral TPJ and anterior temporal lobes. The most notable difference between connectivity clusters 3 and 4 is that only the more dorsally located cluster 3 shows connectivity to lateral prefrontal and premotor areas. All connectivity-based atlases we used here were made only for the right hemisphere, and to our knowledge, no corresponding parcellations exist for the left side. For labeling left hemispheric activations in our meta-analysis, we therefore assumed a homologue parcellation on the left side as for the right side.

## 3. Results

### 3.1. Individual meta-analyses

#### 3.1.1. False belief vs. photo

Neuroimaging studies have followed developmental psychology in using false belief stories as the prototypical problem for theory of mind reasoning (Fletcher et al., 1995; Gallagher et al., 2000). More recently, it has become apparent that the logical structure of false belief stories systematically differs from that of many control stories. This has been circumvented by introducing the so-called “false” photograph control story (Zaitchik, 1990). In this kind of control condition, subjects are required to represent the outdated content of a physical representation such as a photograph or map. We found 15 imaging studies that used the false belief > photo contrast, which had been first published by Saxe and Kanwisher (2003). Examples of this task are provided in Table 1 in the section ‘False belief vs. photo’.

We performed a meta-analysis on the reported activation maps for the contrast false belief vs. photo stories. Results are shown in the sections ‘False belief vs. photo’ in Fig. 2 and Table 2. We found the largest areas of convergence in left and right temporo-parietal cortices. Activation peaks were located in the right posterior superior temporal gyrus corresponding to connectivity cluster TPJp, and the left posterior middle temporal gyrus corresponding to connectivity cluster IPL. Two large areas of convergent activation were also found in the precuneus and in the medial prefrontal cortex.

The posterior area included parts of the bilateral precuneus and posterior cingulate gyrus. The anterior area had its activation peak on the border between mPFC connectivity clusters 3 and 4, and extended to more ventral parts of the mPFC and anterior cingulate gyrus. Finally, an additional area of reliable activation was found in the right anterior temporal lobe and adjacent parts of the insula. Jackknife sensitivity analysis showed that all findings of the meta-analysis were highly reproducible, and remained significant for 15/15 combinations of studies using the leave-one-out method. Heterogeneity analysis found significant between study variance only for the peak in the precuneus, whereas all other findings showed no significant variability across studies.

Eleven studies were excluded from the main analyses for the sake of conceptual homogeneity or for methodological reasons. Scholz et al. (2009) was excluded because only ROI data were reported. Another study mixed judgments about false belief and subjective preference (Jenkins and Mitchell, 2010). Other studies relied on various nonverbal formats to study reasoning about false belief, such as comics, pictures and videos. Some of them contrasted activation for false belief versus true belief reasoning (Hooker et al., 2008; Rothmayr et al., 2011; Sommer et al., 2010; Grèzes et al., 2004a, 2004b, 2006; see also Schuwerk et al., in press), while others contrasted false belief reasoning with reasoning about physical reality (Doehnel et al., 2012; Marjoram et al., 2006; Samson et al., 2004). Unfortunately, nonverbal false belief studies were too few in number and too heterogeneous to form a separate task group. In particular, the contrast false belief versus true belief reasoning would deserve a separate task group because of its conceptual uniqueness.

Another 10 studies used methodologically less stringent control conditions and story-based formats to test false belief. We report them in a separate meta-analysis in Supplementary Materials S2. In general, meta-analytic findings for this task category were less robust compared to findings for the false belief vs. photo. One explanation for these weaker findings may be that heterogeneity and methodological variability for these studies was much larger than for the false belief vs. photo. For example, some studies presented relatively long text passages with plenty of information with no relevance for belief-reasoning (e.g., Fletcher et al., 1995). Including these studies in our main analysis would therefore introduce substantial variance which may not be due to clearly identifiable conceptual differences between tasks, but rather due to haphazard methodological differences.

#### 3.1.2. Trait Judgments

Inspired by former findings of brain areas specialized for conceptual knowledge about different classes of inanimate objects (e.g., tools, houses), early studies presenting trait judgment tasks aimed at finding specialized brain areas for conceptual knowledge about persons (e.g., Mitchell et al., 2002). Since then, trait judgment tasks have been very popular in the field. Common to all studies in this group, the experimental task contains written material which provides information about traits of a person (adjectives, opinions, or personal episodes). The section ‘Trait Judgments’ in Table 1 gives examples. In most of the studies, the described person was not visible; however, three studies presented a photograph of the face of the target person. Different types of control tasks were used in studies of trait judgments. In a number of control tasks attention was diverted away from mental state processing by asking for a lexical judgment on a personality trait word (e.g., *Is this word written in upper- or lower-case?*). Other control tasks contained words or statements which had no mental-state related content.

We performed a meta-analysis on the reported activations for contrasts of trait-diagnostic information processing > no

**Table 1**

Examples from each task-group in our meta-analysis.

Author	Img.	Experimental task	Control task
<b>False belief vs. photo (3 out of 15 studies, for more information see Supplementary Table S1.1)</b>			
Aichhorn (2009)	fMRI n = 21	Read a short vignette involving a person holding a false belief. Predict the behavior of that person based on her belief. e.g., <i>'Julia sees the ice cream van go to the lake. She doesn't see that the van turns off to the town hall. Therefore, Julia will look for the ice cream van at the ...? (lake or town hall)'.</i>	Read a short vignette involving a photograph of the past, and a description how things shown on the photo have changed by now. Answer a question about the outdated scene shown on the photo. e.g., <i>'Julia takes a picture of the ice-van in front of the pond. The ice cream van changes to the market place; the picture gets developed. On the picture, the ice-van is at the ...?' (pond or market place).</i>
Saxe (2003)	fMRI n = 21	Read a short vignette involving a person holding a false belief. Answer a question about her belief. e.g., <i>'John told Emily that he had a Porsche. Actually, his car is a Ford. Emily doesn't know anything about cars so she believed John. When Emily sees John's car, she thinks it is a ...?' (Porsche or Ford).</i>	Read a false-photograph vignette. Answer a question concerning the outdated content in the photo. e.g., <i>'A photograph was taken of an apple hanging on a tree branch. The film took half an hour to develop. In the meantime, a strong wind blew the apple to the ground. The developed photograph shows the apple on the ...? (tree or ground)'.</i>
Lee (2011)	fMRI n = 26	Read a short vignette involving a person holding a false belief. Answer a question about her belief. e.g., <i>'David knows that Ethan is very scared of spiders. Ethan, alone in the attic, sees a shadow move and thinks it is a burglar. David hears Ethan cry for help. David assumes that Ethan thinks he has seen ...?' (a spider or a burglar).</i>	Read a false-photograph vignette. Answer a question concerning the outdated content in the photo. e.g., <i>'Amy made a drawing of a treehouse three years ago. That was before the storm. We built a new treehouse last summer, but we painted it red instead of blue. The treehouse in Amy's drawing is ...? (red or blue)'.</i>
<b>Trait judgments (3 out of 15 studies, see Supplementary Table S1.2)</b>			
Ma (2011)	fMRI n = 30	Read written statements conveying trait diagnostic information about persons (describing behavior). Then read a single trait-adjective and indicate whether it is consistent with the behavior of that person. e.g., <i>'Tolvan gave her sister a hug ... consistent with "friendly"?'</i>	Read written statement about a person doing something. This behavior is neutral and does not convey trait diagnostic information about the person. Indicate the gender of the person in the sentence. e.g., <i>'Tolvan gave her mother a bottle ... is Tolvan male or female?</i>
Mitchell (2002)	fMRI n = 34	Read an adjective. Indicate whether it can be true for a hypothetical person. e.g., <i>"nervous" ... can it be true for "David"?"</i>	Read an adjective. Indicate whether it can be true for an object. e.g., <i>"sundried" ... can it be true for "grape"?</i>
Zhu (2007)	fMRI n = 13	Read a personality trait adjective (e.g., <i>brave, childish</i> ) and indicate if it correctly describes a former American president ( <i>Bill Clinton</i> ).	Read a personality trait adjective (e.g., <i>brave, childish</i> ) and indicate if it is written in lower- or upper-case.
<b>Strategic games (3 out of 9 studies, see Supplementary Table S1.3)</b>			
Gallagher (2002)	PET n = 9	Play "stone, paper, scissors" with a human opponent. Select one option (e.g., <i>stone</i> ). After that, the option of your opponent is shown. The winner gets rewarded (winner is determined by a set of rules, e.g., <i>stone beats scissors</i> ).	Play "stone, paper, scissors" with a computer. You are informed that the computer chooses by a simple algorithm. Select one option.
Kircher (2009)	fMRI n = 14	Play the prisoner's dilemma game (iterated version). You play with a human player for game points. Both players choose a cooperative or defective strategy on each trial. If both players choose defective, they gain almost no game points at all. If both choose cooperative, both gain some game points. If players choose differently, the defective player gains more points.	Play the prisoner's dilemma game (iterated version). You play with a computer.
Sripada (2009)	fMRI n = 26	Play a game like the ultimatum game (iterated version) with a human opponent. Two players must share an amount of money. Player 1 makes an offer how to split the amount, and player 2 can accept (both receive their share) or decline (no one receives money).	Play a game like the ultimatum game (iterated version) with a computer.
<b>Social animations (3 out of 14 studies, see Supplementary Table S1.4)</b>			
Blakemore (2003)	fMRI n = 10	Watch a video animation of two interacting triangles (e.g., <i>surprising one another</i> ). Answer questions concerning the contingency (e.g., <i>is there an intention?</i> ) between movements of the two shapes.	Watch video animation of two interacting triangles (e.g., <i>surprising one another</i> ). Answer questions concerning the physical movement of the first shape (e.g., <i>did the velocity change?</i> ).
Castelli (2000)	PET n = 6	Watch a video animation of two interacting triangles (e.g., <i>mother and child are playing</i> ). Explain verbally what was happening (after fMRI).	Watch video animation of two randomly moving triangles. Explain verbally what was happening (after fMRI).
Martin (2003)	fMRI n = 12	Watch a video animation of simple geometrical shapes depicting a social interaction. Indicate which action was depicted – select a word from several alternatives (e.g., <i>dancing, fishing, sharing...</i> ).	Watch a video animation of simple geometrical shapes depicting a mechanical action. Indicate which action was depicted – select a word from several alternatives (e.g., <i>billiards, bowling...</i> ).
<b>Mind in the eyes (3 out of 10 studies, for more information see Supplementary Table S1.5)</b>			
Baron-Cohen (1999)	fMRI n = 12	View photographs of eyes. Indicate which of two words (e.g., <i>concerned</i> versus <i>unconcerned</i> ) describes the mental state of that person.	View photographs of eyes. Indicate if the person is male or female.
Mitchell (2005b)	fMRI n = 18	View photograph of a face. Indicate how pleased the person was about being pictured.	View Photograph of a face. Indicate how symmetrical the face is.
Platek (2004)	fMRI n = 5	View photographs of eyes. Think about the mental state of the person depicted (no response).	View a fixation cross (no response).

Table 1 (Continued)

Author	Img.	Experimental task	Control task
<b>Rational Actions (3 out of 10 studies, see Supplementary Table S1.6)</b>			
Brunet (2000)	fMRI n = 8	View a cartoon story and predict what will happen based on intentions of a character (no false belief). Choose a logical story ending from several options shown in pictures. e.g., <i>A prisoner is in his cell. First, he breaks the bars of his prison window. Then he walks to his bed. Participants must indicate what will happen next ... the prisoner ties a rope from the sheets on his bed/the prisoner shouts out loud.</i>	View a cartoon story and predict what will happen based on physical causality. Choose a logical story ending from several options shown in pictures. e.g., <i>A person is standing in front of a slide. A large ball is coming down this slide, heading toward the person standing there. Participants must indicate what will happen next ... the ball is knocking over the person/the ball is resting on the ground and the person is standing next to it.</i>
Villareal (2012)	fMRI n = 19	You see a photograph of a person performing a gesture involving a symbolic connotation (e.g., <i>asking for the bill in a restaurant</i> ). Indicate if a gesture was shown and what it was expressing. (e.g., <i>asking for the bill/hitchhiking on the road/nothing</i> ).	You see a photograph of a person simply sitting or standing (e.g., <i>sitting in a restaurant</i> ). Indicate if a gesture was shown and what it was expressing. (e.g., <i>asking for the bill/hitchhiking on the road/nothing</i> ).
Walter (2004) (exp 1)	fMRI n = 13	View a cartoon story showing two persons. One is making a communicative gesture (e.g., <i>is pointing to a bottle to request it</i> ). Choose a logical story ending from three options shown in pictures.	View a cartoon story showing some objects making contact because of physical causality (e.g., <i>a gust of wind blows a ball, so it knocks over several bottles</i> ). Choose a logical story ending from three options shown in pictures.

trait-diagnostic information processing. Sections 'Trait Judgments' in Fig. 2 and Table 2 summarize the results. A medial prefrontal area showed absolute highest convergence, with its peak in mPFC connectivity cluster 3. The area extended ventrally to mPFC connectivity cluster 4, and dorsal-posteriorly to connectivity cluster 2. Another area of convergent activation was found in the posterior cingulate and parts of the precuneus. Convergence was also found in the temporo-parietal cortices, with activation peaks in the right posterior superior temporal gyrus corresponding to connectivity cluster TPJp, and the left posterior middle temporal gyrus corresponding to connectivity cluster IPL. In addition, two areas of convergent activation were found in bilateral anterior temporal lobes. Jackknife sensitivity analysis showed high replicability for bilateral temporo-parietal findings (14/15 combinations left, 15/15 right), medial prefrontal and posterior cingulate findings (15/15 combinations for both peaks). The left anterior temporal finding also showed high replicability (15/15), whereas the right anterior temporal peak could only be reproduced for 11/15 combinations. Heterogeneity analysis found significant between study variance for all left hemispheric areas (left anterior temporal, left temporo-parietal) and for the posterior cingulate, which indicates that these areas may be activated in some trait judgment tasks more strongly than in others. No significant variance was found for the findings in the medial prefrontal, right temporo-parietal and right anterior temporal cortices.

Twelve studies were excluded from the main analyses of trait judgments. Some studies contrasted activation for other versus self-related judgments (Blackwood et al., 2003; D'Argembeau et al., 2007; Gutches et al., 2007; Vanderwal et al., 2008) or used another form of mental state judgment as a comparison condition (Cloutier et al., 2011; Ray et al., 2010; Tamir and Mitchell, 2010). Other studies only reported activation for self-related judgments (Schroeter et al., 2010) or reported activation for a mix between self- and other related judgments (Lombardo et al., 2010; Modinos et al., 2009). Pfeifer et al. (2007) was excluded because this study used passive rest as a control condition. Heberlein and Saxe (2005) presented a kind of social animation and asked for a trait judgment. We excluded this task from our analysis because it could not be clearly assigned to one of our task groups (i.e., trait judgments or social animations). One PET study included in our meta-analysis (Craik et al., 1999) reported extremely high effect-sizes (highest t-values > 100) obtained from a sample of only 6 participants. Because we could not clarify if these effect-sizes were determined by a random effects analysis, we replaced them in our meta-analysis by the software default for missing values.

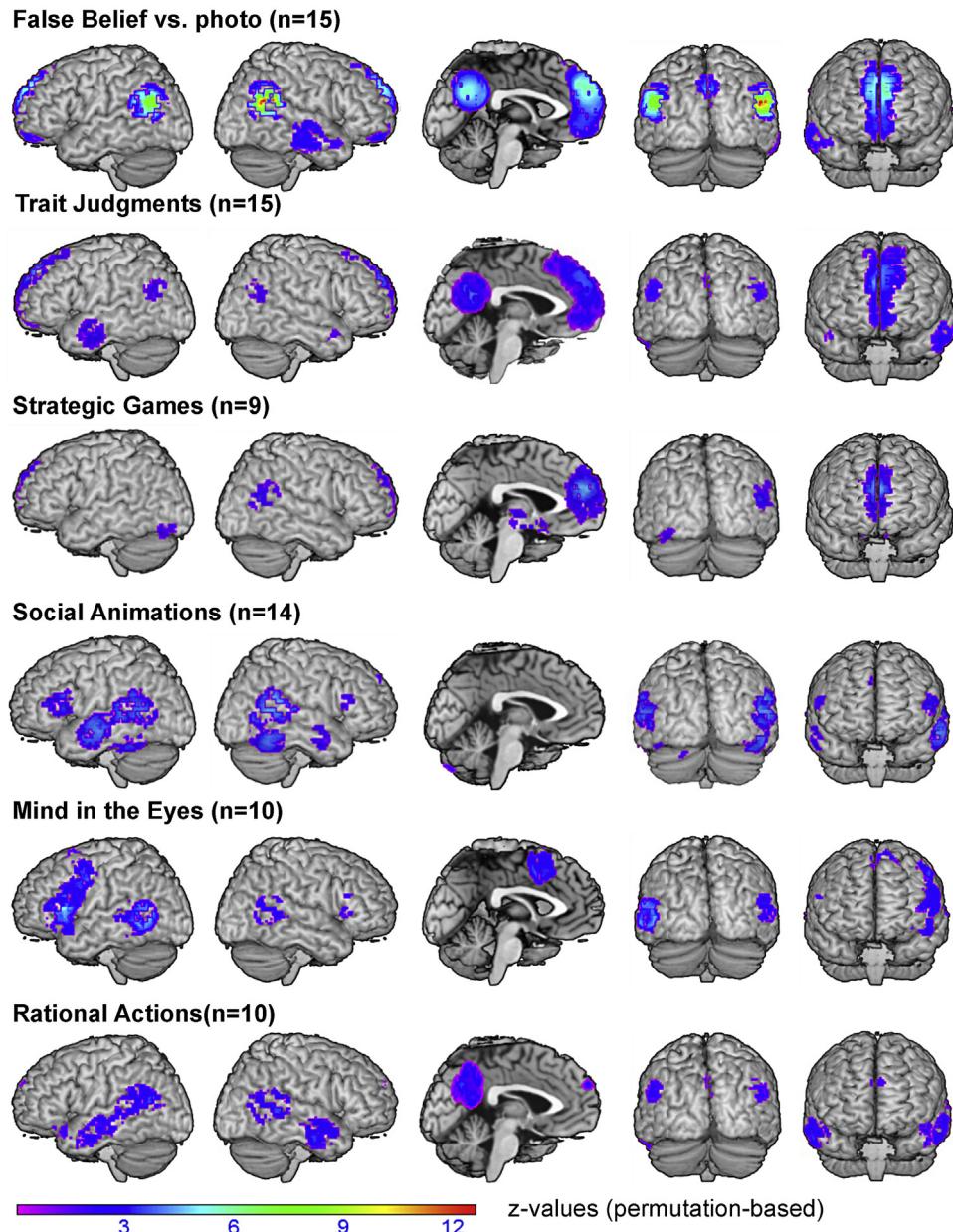
### 3.1.3. Strategic games

Researchers asked participants in the scanner to play a game with another person. The hypothesis was that feedback from a social partner – indicated by her moves in the game – is spontaneously used to infer her intentions, even if participants are not explicitly told to mindread (Rilling et al., 2004). Then, brain activation measured during strategic games was compared with activation found in studies that explicitly asked for a judgment about mental states, such as false belief. Overlaps in activation were taken to support the hypothesis of an "intentional stance" (e.g., Dennett, 1971), i.e., that people have a disposition to reason about the beliefs, desires and intentions of others to predict behavior (Gallagher et al., 2002). In all strategic games reviewed here, participants were asked to play a game where they could compete or cooperate with another player. Section 'Strategic Games' in Table 1 gives examples. Players could not see each other, but they were told about the decision of the other player in the game. The most popular game used by these studies is the classic prisoner's dilemma. In all experimental tasks, the other player was human. In the control tasks, the other player was a computer (e.g., following a simple algorithm).

A meta-analysis was performed on the reported activations for contrasts of playing with a human > playing with a computer. See 'Strategic Games' in Fig. 2 and Table 2 for results. The largest area was found in the mPFC with its peak in mPFC connectivity cluster 3. The area extended to the anterior cingulate gyrus ventrally and the right posterior frontal cortex (mPFC connectivity cluster 5) laterally. Another large area was found in the anterior cingulate gyrus and thalamus. Furthermore, we found activation in a right posterior middle temporal area, with subpeaks in connectivity clusters IPL and TPJa. A smaller area was also found in the left fusiform gyrus. Jackknife sensitivity analyses showed high replicability for all findings (9/9 combinations) except for the fusiform area, where the peak was only replicated in 6 out of 9 combinations. Heterogeneity analysis found no significant between-study variability for the main findings of the meta-analysis of strategic games.

We excluded 7 studies from the meta-analysis of strategic games. Tomlin et al. (2006) was excluded because they only reported ROI data. Bhatt and Camerer (2005) and Krill and Platek (2012) used control conditions which likely engaged reasoning about mental states. Other studies, using low-level control conditions (Decety et al., 2004; Elliott et al., 2006; Rilling et al., 2008), were excluded for the sake of homogeneity. Another group of studies were excluded because they did not use a subtraction design with activation and control tasks, but computational modeling

## Meta-analyses for individual task groups



**Fig. 2.** Results of separate meta-analyses for task groups. Colors represent probability values from statistical permutation testing (z-values). Maps were thresholded at voxel-wise threshold of  $p < .005$  uncorrected and a cluster extent threshold 10 voxels.

approaches (e.g., Behrens et al., 2008; Hampton et al., 2008; Yoshida et al., 2010).

### 3.1.4. Social animations

Social animations were introduced because researchers (Castelli et al., 2000, 2002) sought for a low-level stimulus that provides an alternative to the high-level verbal or cartoon-based materials that predominated the field in the beginning (e.g., Baron-Cohen et al., 1994; Fletcher et al., 1995; Happé et al., 1996). The idea was to investigate whether theory of mind areas found in earlier verbal studies could also be triggered by a very minimalistic and low-level input, which was aimed at showing the existence of a dedicated neurocognitive mechanism for mental-state attribution. Studies in this category presented video animations of simple

geometrical shapes, which were introduced by Heider and Simmel (1944). The section ‘Social animations’ in Table 1 gives examples. The majority of studies in this category used the following experimental paradigm: In the experimental task, a movie showed a number of geometrical shapes (e.g., two triangles), moving across the display. The movements portrayed actions which are typical for an intentional or social interaction. In the control condition, a movie showed geometrical shapes in random or purely mechanical movements (e.g., resembling the movement of billiard balls on the table). For each movie, participants were asked to explain/decide if an interaction was portrayed. A few studies in this category showed similar movies in the activation and control tasks, but used different instructions. In the experimental task, participants were asked to focus on the social interactions shown in the movies. In the control

**Table 2**

Results of meta-analyses for individual task groups.

Label	Cluster center			z-val.	vx	JK	Het	Individual foci			Label
	x	y	z					x	y	z	
<b>False belief vs. photo (15 studies)</b>											
R precuneus (BA 7)	8	-59	35	6.50	1407	15	2.83	4	-42	35	R cingulate
								-1	-49	25	L post. cingulate
								0	-42	35	cingulate
								0	-62	33	precuneus
R sup. temporal (BA 39, TPJp)	56	-56	25	13.22	887	15	n.s.	56	-47	24	R supra./IPL B
								58	-52	42	R inf. par./IPL D, C
								45	-59	39	R angular/IPL D, E
								55	-67	15	R mid. temp./-
R insula (BA 13)	49	-8	-11	3.36	628	15	n.s.	51	-9	-7	R sup. temporal
								61	-16	-15	R mid. temporal
								51	-15	-26	R sub-gyrus
								40	-16	-15	R caudate
L mid. temporal (BA 39, IPL D)	-55	-65	27	8.44	960	15	n.s.	44	-28	-21	R fusiform
								-50	-61	40	L angular/IPL E, D
								-59	-61	45	L inf. par./IPL D, E
								-50	-56	31	L sup. temp./TPJp
L sup. frontal (BA9, cl 3, 4)	-7	58	21	6.29	2095	15	n.s.	9	48	-14	L supra./TPJp
								6	50	31	R ant. cing./-
								-1	51	-7	R med. front./cl 3
								-3	54	28	L ant. cing./-
								6	58	25	R sup. front./cl 3
<b>Trait judgments (15 studies)</b>											
R sup. temporal (BA 39, TPJp)	47	-56	20	2.23	169	14	n.s.	47	-78	34	R mid. temp./-
								54	-68	35	R angular/IPL D
R sup. temporal (BA 38)	44	8	-26	2.00	33	11	n.s.	-1	52	31	L med. front./cl 3
R med. front. (BA 9, cl 3)	6	51	24	4.38	2276	15	n.s.	-1	54	6	L med. front./cl 4
								3	49	-3	R ant. cing./-
								-1	54	6	L ant. cing./cl 4
								-5	33	53	L sup. front./cl 2
								8	33	53	R sup. front./cl 2
L mid. temporal (BA 39, IPL D)	-55	-65	27	2.28	330	15	2.72	-57	-70	26	L mid. temp./-
								-50	-59	25	L supra./TPJp
L post cingulate (BA 31)	-3	-56	26	3.64	970	15	6.26	-11	-65	24	L precuneus
								4	-54	23	R post. cingulate
								4	-53	34	R precuneus
L mid. temporal (BA 21)	-57	-9	-21	2.82	417	15	3.18	-49	-14	-25	L sub-gyrus
								-51	-7	-25	L inf. temporal
								-53	-4	-14	L sup. temporal
<b>Strategic Games (9 studies)</b>											
R mid. temp. (BA 19, -)	48	-60	16	3.14	422	9	n.s.	50	-46	24	R inf. par./TPJp
								58	-62	28	R sup. temp./IPL D
L fusiform (BA 19)	-34	-70	-14	2.14	241	6	n.s.	-36	-84	-14	L inf. occipital
L ant. cing. (-)	-2	2	-4	2.58	657	9	n.s.	-42	-70	-20	L cerebellum
								-10	-20	4	L thalamus
								2	2	-4	R ant. cingulate
								16	-16	0	R thalamus
								12	-2	-6	R glob. pallidus
L med. front. (BA 32, cl 3)	-2	46	26	4.53	1494	9	n.s.	4	46	28	R med. front./cl 3
								-8	40	0	L ant. cing./-
								18	48	32	R sup. front./cl 5
<b>Social animations (14 studies)</b>											
R sup. temporal (BA 13, TPJp)	53	-31	9	4.55	1664	14	n.s.	53	-62	3	R mid. temp./-
								62	-63	-6	R inf. temp./-
								64	-51	24	R supra./TPJp
								51	-53	-16	R fusiform/-
								58	-48	13	R sup. temp./-
								55	-34	2	R sup. temp./-
								40	-33	-4	R caudate/-
								49	-23	10	R trans. temporal/-
R thalamus	12	-13	-1	2.78	98	11	n.s.	8	-23	4	R thalamus
R sup. temporal (BA 21)	57	-4	-16	3.06	141	14	4.54	57	-6	-23	R mid. temporal
R inf. frontal (BA 45)	57	29	14	2.76	54	10	n.s.	57	-5	-30	R mid. temporal
R med. frontal (BA 8, cl 3)	6	59	32	2.66	10	8	n.s.	13	-31	-23	R inf. frontal
L cerebellum	-23	-76	-28	2.79	13	9	n.s.	-49	-53	-14	L fusiform
L cerebellum	-31	-45	-20	4.11	704	14	n.s.	-47	-31	-23	L fusiform
L mid. temporal (BA 21)	-62	-15	-13	4.10	1292	14	n.s.	-55	-66	10	L mid. temp./-
								-65	-45	28	L supra./TPJp

Table 2 (Continued)

Label	Cluster center			z-val.	vx	JK	Het	Individual foci			Label
	x	y	z					x	y	z	
L mid. frontal (BA 46)	-46	23	19	3.29	297	14	n.s.	-59	-51	13	L sup. temp./-
								-48	-37	20	L insula/TPJa
								-61	-34	7	L sup. temp./TPJa
								-60	-10	-9	L mid. temp./-
								-66	-7	-21	L inf. temp./-
								-44	15	8	L insula
								-44	22	7	L inf. frontal
								-57	39	10	L mid. frontal
<b>Mind in the eyes (10 studies)</b>											
R mid. temporal (BA 19, -)	56	-63	16	3.31	311	9	n.s.	55	-64	3	R mid. temp./-
								58	-60	21	R sup. temp./TPJp
								47	-36	9	R sup. temp./TPJa
R inf. frontal (BA 45)	47	22	6	2.79	28	8	n.s.	42	19	0	R insula
R inf. frontal (BA 9)	60	25	19	2.38	11	5	n.s.	-55	-66	12	L mid. temp./-
L mid. temporal (BA 37, -)	-51	-62	5	4.84	655	10	n.s.	-63	-57	16	L sup. temp./TPJp
								-53	-53	-12	L inf. temp./-
								-51	-43	-13	L fusiform/-
								-44	-32	8	L sup. temp./-
L cingulate gyrus (BA 24, -)	-5	8	42	3.75	655	10	n.s.	-3	6	49	L med. front./-
								8	7	51	R med. front./cl 1
								-13	14	59	L sup. front./cl 1,2
								6	17	47	R sup. front./-
								-9	27	40	L cing./-
L inf. frontal (BA 45)	-46	24	7	5.27	1465	10	n.s.	-37	3	38	L precentral
								-48	11	51	L mid. frontal
								-55	7	11	L inf. frontal
								-55	16	17	L inf. frontal
								-42	17	8	L insula
								-46	26	27	L mid. frontal
								-51	37	-12	L inf. frontal
<b>Rational actions (10 studies)</b>											
R precuneus (BA 7)	6	-48	38	3.64	894	10	n.s.	-5	-60	31	L precuneus
								6	-59	10	R post. cingulate
								10	-53	32	R cingulate
								-3	-50	18	L post. cingulate
R mid. temporal (BA 21, -)	60	-45	6	3.57	593	10	n.s.	53	-68	12	R mid. temp./-
								53	-53	11	R sup. temp./-
								58	-41	19	R sup. temp./TPJa
R mid. temporal (BA 21)	59	-11	-25	3.54	545	10	n.s.	64	-38	3	R mid. temp./-
								64	-20	-10	R mid. temporal
								57	-12	-36	R inf. temporal
								57	1	-15	R sup. temporal
								48	6	-33	R mid. temporal
L mid. temporal (BA 39, IPL E)	-50	-65	23	3.74	1242	10	n.s.	-55	-49	35	L supra./TPJp
								-50	-56	29	L sup. temp./IPL C, E
								-61	-34	9	L sup. temp./TPJa
								-59	-16	-2	L sup. temp./-
								-55	-4	-14	L mid. temp./-
L sup. frontal (BA 9, cl 4)	-1	62	23	2.16	31	7	n.s.				

JK, number of subsamples in jackknife analysis that could reproduce finding.

Het, significance of between-study heterogeneity test.

task, participants had to focus on the mechanical properties of the geometrical shapes (e.g., indicate if the velocity of the movement had changed).

The meta-analysis for this category was based on contrasts between social or intentional interactions > physical movements. Results are summarized in the sections 'Social Animations' in Fig. 2 and Table 2. We found the largest areas of convergence in bilateral temporo-parietal cortices. The right area had its activation peak in connectivity cluster TPJa in the connectivity-based parcellation scheme, and also covered parts of connectivity cluster TPJp. On the left side, activations were only found in connectivity cluster TPJa. Other areas of convergence were found in the right anterior temporal lobe, left cerebellum, and left middle/inferior frontal gyrus. Smaller areas of convergence were found in the right thalamus, right inferior frontal gyrus and in medial prefrontal cortex (connectivity cluster 3). Jackknife sensitivity analyses showed

high reproducibility for main findings in bilateral temporo-parietal areas, the right anterior temporal lobe, the left middle/inferior frontal gyrus and the left cerebellum (significance remained in 14/14 combinations). Lower replicability was found for findings in the right thalamus (11/14), right inferior frontal gyrus (10/14) and the right medial prefrontal cortex (8/14). Heterogeneity analysis found significant between-study variance for the findings in the right anterior temporal lobe and in the right inferior frontal gyrus. We excluded five studies from the meta-analysis of social animations. Two studies were excluded because they did not report activation coordinates (Scheibel et al., 2011; Schultz et al., 2003). Three studies used animations which portrayed only causal relations between mechanically moving objects, but no social content (Blakemore et al., 2001; Fonlupt, 2003; Straube et al., 2011). Because of a similar issue as described in the previous section Trait Judgments, we replaced extremely high effect sizes reported in one

small-sampled study (Castelli et al., 2000) by the software default for missing values.

### 3.1.5. Mind in the eyes

The mind in the eyes task was used in fMRI as an attempt to dissociate brain mechanisms subserving general intelligence from those dedicated to social intelligence (Baron-Cohen et al., 1999). This was based on findings that adults with high-functioning autism (Baron-Cohen et al., 1997) as well as parents of children with autism (Baron-Cohen and Hammer, 1997) showed deficits on the mind in the eyes task, but not children with William's syndrome (Tager-Flusberg et al., 1998). Hence, the task was considered as an advanced theory of mind test involving mind-reading (Baron-Cohen et al., 1999). Examples for the task are provided in Table 1. Almost all studies in this category used the following experimental paradigm: In the experimental task, participants saw a photograph of a pair of eyes. In addition, two mental state adjectives were shown. Participants were asked to indicate which mental-state adjective best describes the expression in the eyes on the photo. The control task also presented a pair of eyes, however, participants were simply asked to indicate the gender or age of the person displayed. A few studies falling into this category showed a complete face on the photo, or a video of a pair of eyes. Again, a mental state judgment was compared to a judgment about physical properties in these studies.

A meta-analysis was performed on the reported activations for contrasts of mental state judgments > physical judgments (e.g., gender, age), see sections 'Mind in the Eyes' in Fig. 2 and Table 2. A left inferior frontal area, with a peak BA 45, showed the highest activation in the map. The area also contained parts of left precentral and middle frontal gyri, as well as in left insula. In the right hemisphere, two smaller lateral prefrontal areas were found in the inferior frontal gyrus pars triangularis (BAs 9 and 45). Further areas of convergent activation across studies were found in bilateral posterior temporal cortices. The left temporal area was substantially larger compared to the right. Both areas had their activation peaks in the posterior middle temporal gyri, and covered connectivity clusters TPJp (left and right side) and TPJa (right side). The left temporal area further extended to the inferior temporal and fusiform gyri ventrally. Finally, an area was found in the middle cingulate gyrus, corresponding to mPFC connectivity clusters 1 and cluster 2 (pre-SMA). Jackknife-sensitivity analysis found perfect replicability for the left temporo-parietal, left middle cingulate and left inferior frontal findings, and good replicability (9/10 combinations remained significant) for the right temporo-parietal finding. Lower replicability was found for the right inferior frontal findings (8/10 and 5/10). Heterogeneity analysis found no significant between-study variance for the findings of the meta-analysis on mind in the eyes tasks.

Four other studies found in the literature were not included in our meta-analysis. Péron et al. (2010) and Baglio et al. (2012) only reported data for patient populations. Mier et al. (2010) and Schulte-Rüther et al. (2011) were excluded because they did not ask for mental states but for basic emotions in their judgments.

### 3.1.6. Rational actions

The initial idea of studies falling into the category rational actions was to produce a non-verbal alternative to false belief stories, which also requires attributing intentions to the protagonist of a story. Another difference is that tasks are not about false belief but rational action goals. Similar to the idea behind the social animations task, comic-strips were used to see whether brain areas dedicated to mental-state attribution responded irrespective of stimulus input (Brunet et al., 2000). Moreover, the non-verbal format was used in imaging studies of theory of mind in schizophrenia, as it circumvents problems linked to speech disorganization in

some patients (Brunet et al., 2003; Sarfati et al., 1997). All tasks in our meta-analysis had non-verbal material. Examples for tasks are given in Table 1, section 'Rational actions'. In the experimental tasks, participants were asked a question about the goal of these actions, e.g., to predict the likely outcome. The correct answer corresponded to the rational (i.e., ordinary) goal pursued by these simple actions. In the control tasks, questions about non-mental aspects of the scenes were asked (e.g., about physical causality).

We performed a meta-analysis of the reported activations for contrasts between inferring action goals > inferring physical causality. Results can be found in the sections 'Rational actions' of Fig. 2 and Table 2. Large areas of convergence were found in bilateral temporo-parietal cortices, with activation peaks in connectivity clusters IPL, TPJp and TPJa on the left side, and TPJa on the right side. Other large areas of convergence were in the precuneus/posterior cingulate and in right anterior temporal lobe. In addition, we found a small area of convergence in mPFC connectivity cluster 4. Jackknife sensitivity analysis showed high replicability for bilateral temporo-parietal, right anterior temporal and precuneus findings (10/10 combinations remained significant), and lower replicability for the finding in the medial prefrontal cortex (7/10). Heterogeneity analysis found no significant between-study variance for rational actions.

We did not include 5 other studies found in the literature. Two studies (Brunet et al., 2003; Ciaramidaro et al., 2007) reported data that had already been used in other studies in our meta-analysis (so there was a problem of dependent samples). Brüne et al. (2011) only reported data for a patient population. German et al. (2004) and Saxe et al. (2004a, 2004b) only asked participants to passively watch human actions, and did not ask about goals or end-states.

### 3.1.7. Summarizing convergent activations across task groups

The first strategy we used to summarize the meta-analytic results across task groups was a pooled meta-analysis over all studies in our sample. This corresponds to the approach taken by previous voxel-based meta-analyses of theory of mind. Table 3 and Fig. 3A show results of the pooled meta-analysis. Consistent with previous work (e.g., Bzdok et al., 2012; Mar, 2011) we found convergent activation along the entire stretch of the bilateral temporal lobes and parts of the bilateral inferior parietal lobuli. Activation peaks were located in the left and right superior temporal gyri corresponding to connectivity clusters TPJp. We found further areas in the bilateral superior temporal gyri corresponding to connectivity clusters TPJa, and in the bilateral angular gyri corresponding to connectivity clusters IPL. In addition, we also found large areas of convergent activation in the precuneus and in the mPFC. The precuneus showed activation in both hemispheres and had its peak on the right side. The frontal area had its peak in mPFC connectivity cluster 3 and extended ventrally to connectivity cluster 4 and dorsal-posteriorly to connectivity cluster 2/pre-SMA. Furthermore, relatively small areas of convergence were found in the left inferior frontal gyrus and in the left fusiform gyrus. Also the right inferior frontal gyrus showed convergent activation, which was subsumed under the large right temporal area.

Jackknife sensitivity analysis showed that all of the main findings (L and R TPJp, mPFC connectivity cluster 3, precuneus, left inferior frontal gyrus) remained significant in 73 out of 73 combinations of studies using the leave-one-out method. Only the peak in the left fusiform gyrus showed less replicability (remained significant in 63 out of 73 analyses). In contrast, heterogeneity analysis showed significant variance across studies for all of the main findings. High and significant heterogeneity was found in the precuneus ( $z = 9.08$ ), left ( $z = 5.82$ ) and right ( $z = 8.83$ ) connectivity cluster TPJp, and in mPFC connectivity cluster 3 ( $z = 5.29$ ).

Next, we characterized the overlap between task groups by means of conjunction analysis. As described in the methods section,

**Table 3**

Results of pooled meta-analysis.

Label	Cluster center			z-val.	vx	JK	Het	Individual foci			Label
	x	y	z					x	y	z	
R precuneus (BA 7)	<b>4</b>	<b>-55</b>	<b>34</b>	6.58	1171	73	9.08	15	-68	29	R precuneus
								-9	-68	29	L precuneus
								0	-55	34	R precuneus
								15	-53	36	R precuneus
R sup. temp. (BA 22, TPJp)	<b>56</b>	<b>-56</b>	<b>18</b>	7.96	2446	73	8.83	56	-72	31	R mid. temp./-
								<b>47</b>	<b>-61</b>	<b>39</b>	R angular/IPL E, D
								<b>64</b>	<b>-52</b>	<b>-3</b>	R mid. temp./-
								56	-47	24	R supramarg./IPL B
								<b>47</b>	<b>-35</b>	<b>12</b>	R sup. temp./TPJa
								40	-35	12	R insula/-
								57	-28	-21	R inf. temp./-
								<b>62</b>	<b>-22</b>	<b>-10</b>	R mid. temp./-
								57	-9	-23	R mid. temp./-
								<b>53</b>	<b>0</b>	<b>-21</b>	R sup. temp./-
								<b>49</b>	<b>24</b>	<b>6</b>	R inf. front.
								55	21	-1	R precentral
L sup. temp. (BA 22, IPL D/TPJp)	<b>-53</b>	<b>-59</b>	<b>20</b>	7.21	1980	73	5.82	-50	-77	29	L mid. temp./-
								<b>-46</b>	<b>-63</b>	<b>41</b>	L angular/IPL E, D
								<b>-66</b>	<b>-52</b>	<b>2</b>	L mid. temp./-
								-52	-47	26	L supramarg./-
								-57	-35	20	L insula/IPL B
								<b>-53</b>	<b>-30</b>	<b>10</b>	L sup. temp./TPJa
								<b>-59</b>	<b>-23</b>	<b>-8</b>	L mid. temp./-
								-60	-15	-13	L mid. temp./-
								<b>-51</b>	<b>0</b>	<b>-19</b>	L mid. temp./-
L fusiform (BA 37)	-51	-46	-8	3.36	14	63	n.s.				
L inf. frontal (BA 13)	-44	27	0	4.40	389	73	0.10	-44	15	8	L insula
								-48	21	19	L inf. frontal
								<b>-46</b>	<b>22</b>	<b>8</b>	L inf. frontal
L sup. frontal (BA 9, cl 3)	<b>-1</b>	<b>56</b>	<b>24</b>	7.74	2295	73	5.29	-16	56	26	L sup. frontal/cl 3
								<b>3</b>	<b>51</b>	<b>-7</b>	R ant. cingulate/cl 4
								-20	48	36	L sup. front./-
								-7	43	52	L sup. front./cl 3
								<b>6</b>	<b>26</b>	<b>55</b>	R sup. front./cl 2
								-5	17	52	L sup. front./cl 2

JK, number of subsamples in jackknife analysis that could reproduce finding.

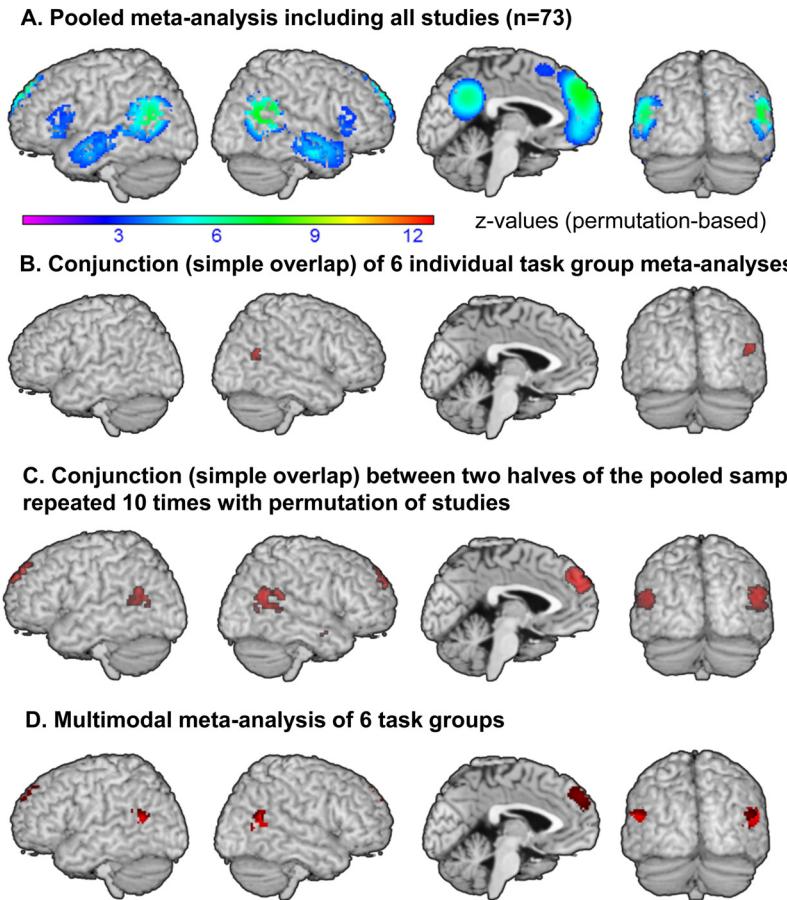
Het, significance of between-study heterogeneity test; Regions highlighted in bold entered our ROI analysis.

**Table 4**

Results of conjunction and overlap analyses.

#	Label	Cluster center			BA	z-val.	vx	Individual foci			Label
		x	y	z				x	y	z	
<b>Conjunction (simple overlap) of 6 individual task group meta-analyses</b>											
1	R mid. temp./TPJp	49	-56	19	21	2.15	69	49	-61	20	R mid. temp./IPL E
								44	-55	18	R mid. temp./-
								56	-55	23	R sup. temp./TPJp
<b>Conjunction (simple overlap) between two halves of the pooled sample—repeated 10 times with permutation of studies</b>											
1	R mid. temp./IPLD	51	-60	20	39	3.66	327	55	-68	12	R mid. temp./-
								48	-58	30	R angular/TPJp
								64	-46	8	R mid. temp./-
								60	-43	26	R supramarg./TPJa
2	L mid. temp./-	-55	-59	19	21	3.58	162	-54	-65	11	L mid. temp./-
								-59	-53	25	L supramarg./TPJp
								-57	-46	14	L mid. temp./-
3	L med. front./-	-1	54	25	10	3.64	257	-3	51	40	L med. front./cl3
								-7	60	37	L med. front./cl3
								-2	61	20	L med. front./cl4
<b>Multimodal meta-analysis of 6 task groups</b>											
1	R sup. temp./TPJp	62	-58	20	39	2.78	184	58	-66	10	R mid. temp./-
								55	-61	16	R mid. temp./-
								56	-55	27	R sup. temp./TPJp
2	L sup. temp./TPJp	-55	-59	20	22	2.59	76	-50	-56	26	L sup. temp./TPJp
								-52	-47	26	L supram./TPJp
3	L med. front./cl 3	-1	54	33	8	1.48	156	-3	44	35	L med. front./cl3
								-4	50	39	L med. front./cl3
								-2	62	24	L med. front./cl4

## Summarizing across task groups



**Fig. 3.** (A) Results of a pooled meta-analysis including all studies from every task group. (B) Results of a conservative conjunction analysis of the six separate meta-analyses. (C) Results of a statistically powerful permutation-based overlap analysis of the six separate meta-analyses. (D) Results of a more liberal multimodal meta-analysis of the six meta-analyses. All results were thresholded at voxel-wise threshold of  $p < .005$  uncorrected and a cluster extent threshold 10 voxels.

we followed three strategies to characterize the overlap between findings for our task groups. First we calculated a simple conjunction overlap between significant findings for the six individual task groups (Fig. 3B). This procedure shows findings which are individually significant in all six meta-analyses at our default threshold of  $p < .005$  uncorrected,  $z > 1$ , and minimum cluster extent 10 voxels. We found only one cluster of consistent activation in our conjunction analysis, which was centered in the right posterior middle temporal gyrus corresponding to connectivity cluster TPJp. This area comprised 69 voxels and extended to parts of the posterior superior temporal gyrus.

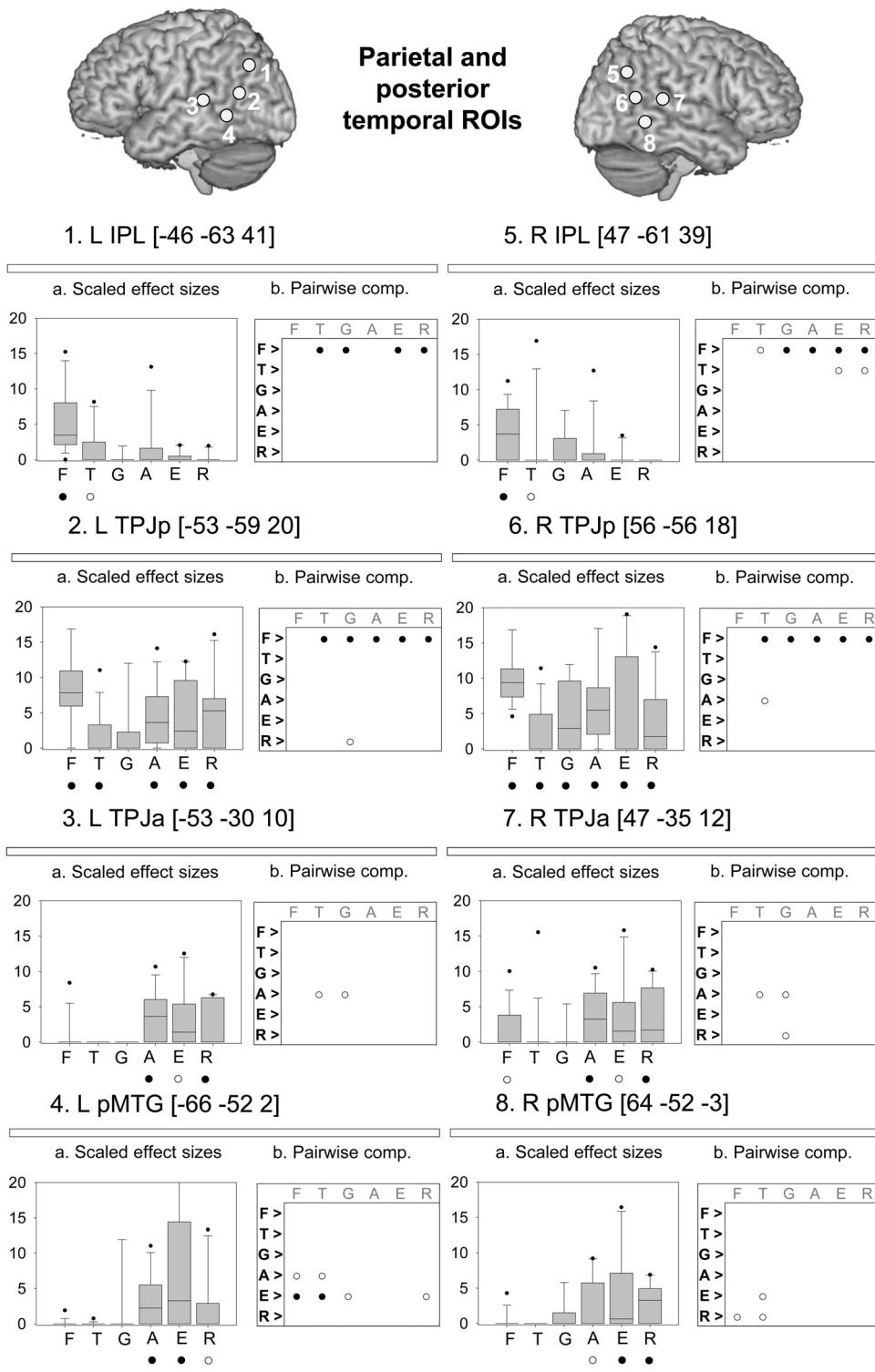
Our second strategy of calculating the overlap between task groups was a permutation-based overlap analysis, repeated ten times for different permutations. Conjunction analysis of all ten overlap images showed three areas of consistent activation. Results are reported in Fig. 3C and Table 4. The largest area of conjoint activation was found in the right posterior middle temporal gyrus with its peak in connectivity cluster IPL, located at 5 mm distance to connectivity cluster TPJp found for our simple conjunction analysis. On the left side, we found a smaller area of conjoint activation, with its peak in the posterior middle temporal gyrus. In addition, we found an area of conjoint activation in the medial prefrontal cortex, covering parts of mPFC connectivity clusters 3 and 4.

Finally, we followed a third strategy to determine the overlap between task groups. We used the statistically more sensitive

multimodal meta-analysis (Radua et al., 2013) to determine areas commonly engaged by the six task groups. Similar to what we found for the permutation-based approach, we found three areas of conjoint activation for the multimodal analysis (see Table 4). Again, the largest area was found in the right temporo-parietal cortex. Its peak was in the right posterior superior temporal gyrus corresponding to connectivity cluster TPJp, at a distance of 13 mm to the peak found for simple conjunction overlap. On the left side, we found a smaller cluster located in connectivity cluster TPJp. Finally, the multimodal analysis found a cluster in the medial prefrontal cortex covering mPFC connectivity clusters 3 and 4.

### 3.2. Region of interest (ROI) analyses

To characterize task-related differences in brain activation in more detail, we relied on a ROI analysis. We extracted for each study an effect-size estimate (Hedges  $g$ ) for each ROI. Estimates are displayed as boxplots for separate task groups in Figs. 4–6. Below these boxplots, we indicated with a full circle if activation in that ROI was found significant in the mean analysis, at a threshold of  $p < .005$ ,  $z > 1$  (determined by permutation tests, see Section 2). Empty circles indicate a trend toward an effect-size above chance level at  $p < .05$ ,  $z > 1$ . Besides boxplots, we also present tables that show the results of pairwise comparisons between task groups for a ROI. Pairwise comparisons were calculated with a linear model



**Fig. 4.** Regions of interest in posterior temporal and parietal brain areas. (a) Box-plots (median; 25th and 75th percentiles; 5th and 95th percentiles) show the distributions of effect-sizes for the studies in each group. Effect-sizes were weighted by intra-study variances. Significant convergence of effect-sizes above zero was determined by randomization tests with SDM; full circles indicate  $p < .005$  uncorrected,  $z > 1$ . Empty circles indicate  $p < .05$ ,  $z > 1$ . (b) Results of pairwise meta-analytic comparisons between all task groups for the regions of interest. Task groups are abbreviated by capital letters. Pairwise comparisons were calculated by linear contrasts with SDM. Full circles indicate a significant difference between two task groups at  $p < .005$ , Bonferroni corrected for the number of pairwise comparisons ( $n = 15$ ). Empty circles indicate  $p < .005$  uncorrected.

that takes into account differences in sample size and variability between meta-analyses. In the tables, full-circles indicate a group-difference at  $p < .005$  Bonferroni corrected,  $z > 1$  (corresponding to an uncorrected  $p < .00033$ ). Empty circles indicate a trend toward a group-difference at  $p < .005$  uncorrected,  $z > 1$ .

### 3.2.1. Temporo-parietal ROIs

We selected four temporo-parietal ROIs in each hemisphere, based on activation peaks from clusters found in the pooled-meta analysis (indicated in boldface in Table 3). In addition to the left and right temporo-parietal cluster peak, we selected three subpeaks to characterize the extent of this area (for details, see Section 2). Our procedure resulted in bilateral ROIs in the anterior and posterior connectivity clusters of the TPJ: Connectivity-cluster TPJa (left:  $x = -53, y = -30, z = 10$ ; right:  $x = 47, y = -35, z = 12$ ) and connectivity cluster TPJp (left:  $x = -53, y = -59, z = 20$ ; right:  $x = 56, y = -56, z = 18$ ). In addition, we selected ROIs in connectivity cluster IPL which is distinct from the TPJ clusters (left:  $x = -46, y = -63, z = 41$ ; right:  $x = 47, y = -61, z = 39$ ). Finally, we selected a posterior middle temporal area, which was located substantially ventral to the TPJ areas (left:  $x = -66, y = -52, z = 2$ ; right:  $x = 64, y = -52, z = -3$ ).

As shown in Fig. 4, ROIs #4 and #8 in the left and right posterior middle temporal gyri showed virtually no activations for the task groups false belief vs. photo, strategic games and trait judgments. A common element in these tasks is that they do not show any behavioral cues of agency or intentionality (e.g., movement, facial expression, gaze). For the other task groups, we found activation in these ROIs. In the left posterior middle temporal ROI #4, we found significant activation for social animations and mind in the eyes ( $zs > 2.7$ ), and a trend for rational actions. Pairwise comparisons found that mind in the eyes showed significantly stronger activation than false belief vs. photo and trait judgments ( $zs > 4.1$ ). In the right posterior middle temporal ROI #8, we found significant activation for mind in the eyes and rational actions ( $zs > 2.3$ ) and a tendency for social animations. Pairwise comparisons showed group differences only as tendencies.

ROIs #3 and #7 in connectivity clusters TPJa showed a related pattern to what we found for posterior middle temporal ROIs. Significant activation was found for social animations and rational actions ( $zs > 2$ ) and tendentially for mind in the eyes on the left side. On the right side, significant activation was again found for social animations and rational actions ( $zs > 2.9$ ), and tendencies were found for mind in the eyes and false belief vs. photo. Only tendencies and no significant group differences were found in bilateral connectivity clusters TPJa. A strikingly different pattern of activations was found for ROIs #2 and #6 in connectivity clusters TPJp. Significant activation was found for all task groups on the right side ( $zs > 2.3$ ). Moreover, pairwise comparisons showed that activation for false belief vs. photo was higher in this area compared to all other task groups ( $zs > 2.5$ ). On the left side, significant activation was found for all task groups but strategic games ( $zs > 1.9$ ). Again, false belief vs. photo activated the ROI stronger than all other task groups ( $zs > 1.9$ ). In left and right connectivity cluster IPL ROI #1 and #5, significant activation was only found for false belief vs. photo ( $zs > 3.6$ ). In addition, trait judgments activated bilateral connectivity cluster IPL ROIs tendentially. Similar to connectivity cluster TPJp, pairwise comparisons also showed for connectivity cluster IPL ROIs stronger activation for false belief vs. photo compared to a number of other task groups. These were trait judgments, strategic games, mind in the eyes and rational actions on the left side ( $zs > 2.2$ ), and strategic games, social animations, mind in the eyes and rational actions on the right side ( $zs > 1.6$ ).

### 3.2.2. Cortical midline ROIs

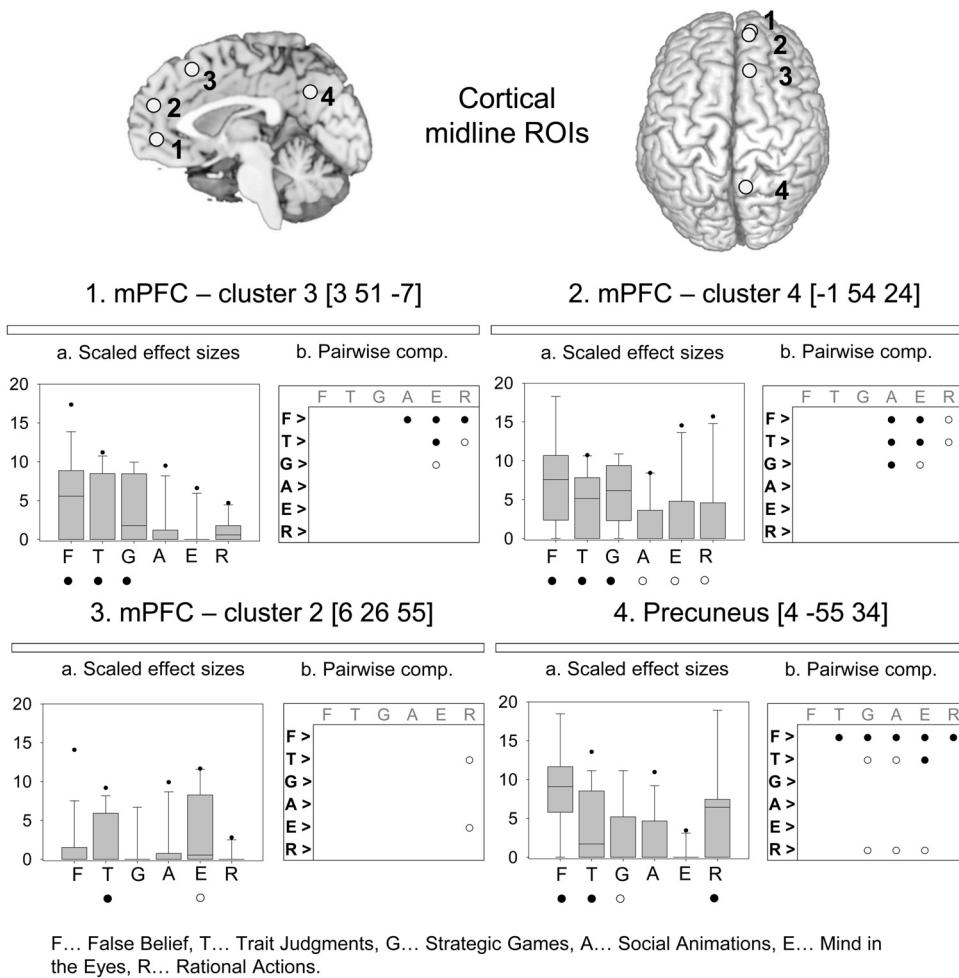
Cortical midline ROIs were selected in a similar fashion to ROIs in the temporo-parietal cortex. We selected the cluster peaks in the medial prefrontal cortex and in the precuneus (see Table 3). In addition, we selected two more local subpeaks in the medial prefrontal cortex, which were (i) at least 20 mm apart from other ROIs and (ii) fell into different connectivity clusters according to the scheme by Sallet et al. (2013). The ROI for the medial prefrontal peak was located in connectivity cluster 4, at  $x = -1, y = 54, z = 24$ . More ventrally, a ROI was located at a subpeak within connectivity cluster 3, at  $x = 3, y = 51, z = -7$ . Finally, a third medial prefrontal ROI was located substantially more dorsal/posterior in connectivity cluster 2/preSMA, at  $x = 6, y = 26, z = 55$ . The ROI in the precuneus was located on the peak at  $x = 4, y = -55, z = 34$ . No sub-peaks were found at a distance of more than 20 mm. Effect-sizes of these ROIs are shown in Fig. 5.

ROI #1 in mPFC connectivity cluster 3/ventral mPFC showed significant activation for false belief vs. photo, trait judgments and strategic games ( $zs > 2.3$ ). Pairwise comparisons found stronger activation for false belief vs. photo compared to social animations, mind in the eyes and rational actions ( $zs > 1.8$ ), as well as stronger activation for trait judgments compared to mind in the eyes ( $z = 1.5$ ). For ROI #2 in mPFC connectivity cluster 4/dorsal mPFC, we again only found significant activation for false belief vs. photo, trait judgments and strategic games ( $zs > 3.5$ ). We also found non-significant tendencies for activation for all other task groups. In ROI #3 in mPFC connectivity cluster 2/pre-SMA, we found significant activation for trait judgments ( $z = 2.1$ ) and a trend for mind in the eyes. Group differences were only found in the form of non-significant trends. Finally, in the precuneus ROI #4, we found significant activation for false belief vs. photo, trait judgments and rational actions ( $zs > 3.0$ ), as well as a tendency for strategic games. Pairwise comparisons showed, similar to what we found for the right TPJp ROI, that false belief vs. photo activated the precuneus ROI #4 significantly stronger than all other task groups ( $zs > 1.6$ ). In addition, trait judgments activated the precuneus ROI #4 more strongly than mind in the eyes ( $z = 2.1$ ) and tendentially also more strongly than strategic games and social animations.

### 3.2.3. Fronto-temporal ROIs

In accordance with findings from previous literature reviews (Gallagher and Frith, 2003; Mar, 2011), our pooled meta-analysis also found local subpeaks in more anterior parts of the bilateral temporal lobes and in the bilateral inferior frontal gyri (see Table 3). Peaks in the anterior temporal lobes were found near the temporal poles (about 7 mm distance in both hemispheres). Coordinates of the anterior temporal ROIs were  $x = -51, y = 0, z = -19$ , and  $x = 53, y = 0, z = -21$ , respectively. We also extracted effect-size estimates for areas in more central parts of the temporal lobes, located about 20 mm posteriorly to anterior temporal ROIs at  $x = -59, y = -23, z = -8$ , and  $x = 62, y = -22, z = -10$ . For the inferior frontal gyri, we selected ROIs at  $x = -46, y = 22, z = 8$ , and  $x = 44, y = 20, z = 12$ .

As shown in Fig. 6, ROIs #1 and #4 in the left and right middle temporal gyri showed activation for social animations and rational actions on the left side ( $zs > 2.4$ ), and for false belief vs. photo and rational actions ( $zs > 2.0$ ) on the right side. In addition, a tendency for activation was also found for social animations on the right side. Pairwise comparisons found stronger activation for social animations and for rational actions compared to strategic games on the left side ( $zs > 2.2$ ), and stronger activation for false belief vs. photo compared to trait judgments and mind in the eyes on the right side ( $zs > 2.1$ ). For ROIs #2 and #5 in the anterior temporal lobes, we found significant activation for trait judgments, social animations and rational actions on the left side ( $zs > 2.4$ ), and significant activation for false belief vs. photo, social animation and rational actions on the right side ( $zs > 2.3$ ). On the left side, we further found stronger



**Fig. 5.** Regions of interest in cortical midline areas. (a) Box-plots show the distributions of effect-sizes; full circles indicate  $p < .005$  uncorrected,  $z > 1$ . (b) Results of pairwise meta-analytic comparisons; full circles indicate  $p < .005$ , Bonferroni corrected. Details same as for Fig. 4.

activation for trait judgments compared to strategic games ( $z = 1.8$ ), whereas no significant group differences were found on the left side. For the left inferior frontal ROI #3, we found significant activation for social animations and mind in the eyes ( $zs > 2.9$ ). Pairwise comparisons found stronger activations for mind in the eyes compared to false belief vs. photo, trait judgments, strategic games and rational actions ( $zs > 3.0$ ). For social animations, we found stronger activation compared to false belief vs. photo ( $z = 3.1$ ), and in trend also compared to other tasks in the left inferior frontal ROI #3. For the right inferior frontal ROI #6, we again found significant activation for mind in the eyes ( $z = 2.4$ ), and activation trends for trait judgments and social animations. Group differences only emerged as non-significant trends.

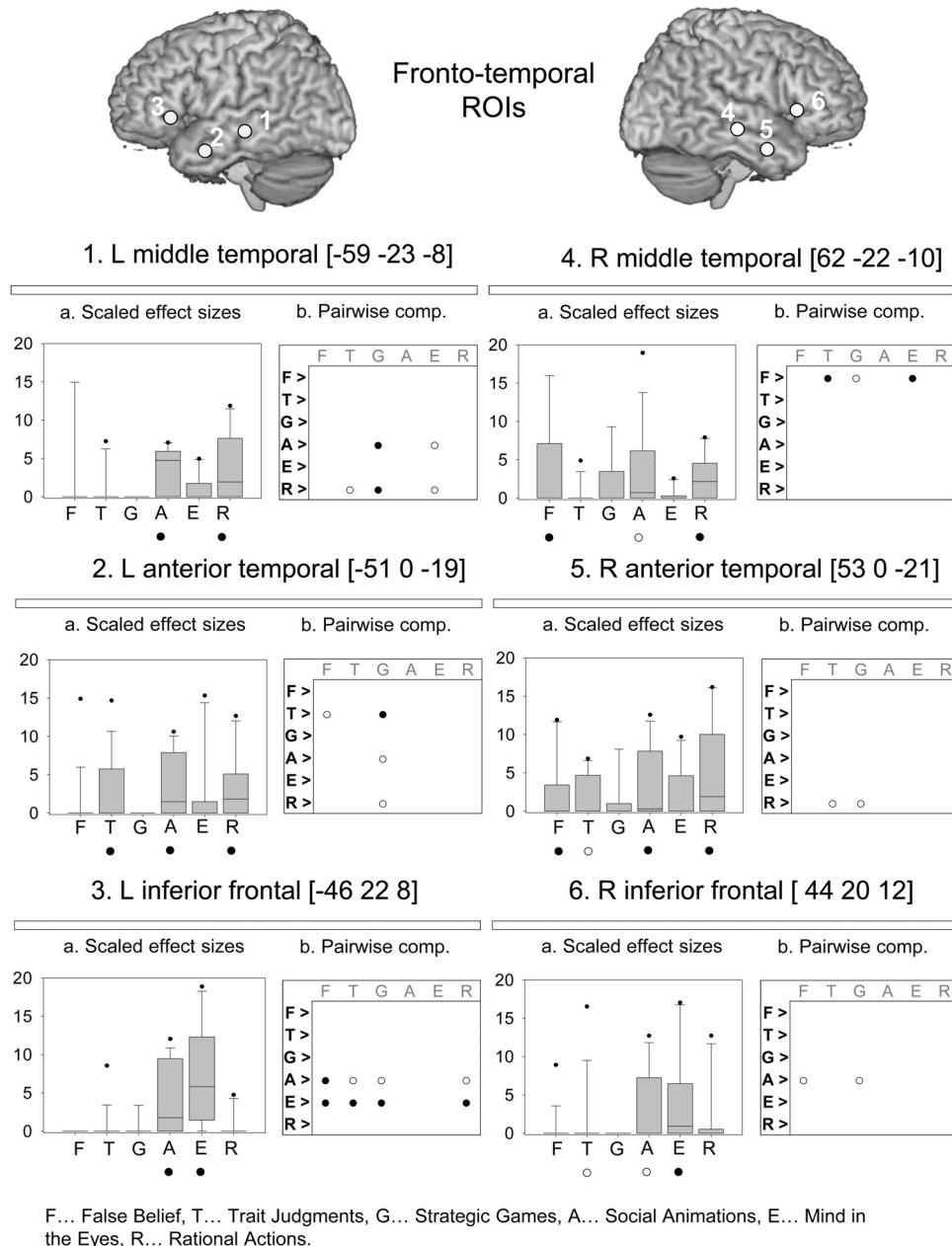
#### 4. Discussion

##### 4.1. The role of conceptual variance for meta-analysing theory of mind

Meta-analyses are a useful way to provide objective summaries of the rapidly growing amount of research in the neuroimaging domain. For a large and heterogeneous field such as theory of mind research, it is important to summarize findings with the appropriate level of detail. Different levels can be chosen, and each has its own merits and drawbacks. The most simple and straightforward way to meta-analyse findings is to put together all studies that call themselves a theory of mind task into a pooled analysis.

This approach was taken by previous meta-analyses on theory of mind (Decety and Lamm, 2007; Bzdok et al., 2012). A clear benefit of the pooled approach is that it has the best statistical power to detect activations, as it is based on a large sample of all available studies. It will show which activations are robust despite all the variability in the field. Large variability, however, is also a drawback of the pooled approach. A pooled meta-analysis puts together studies using different stimulus materials and control conditions and studies with variable methodological quality. This inevitably will create a lot of noise. To check for variability in our pooled analysis, we performed heterogeneity tests for the main findings. In support of our expectation, we found that all main findings (TPJ, mPFC, precuneus) were accompanied by significant inter-study heterogeneity, which means that the observed between-study variance was larger than that resulting from sampling error alone. The presence of high variability in our pooled analysis indicates that this approach may be missing some details about the functional neuroanatomy of theory of mind. Another potential problem of a pooled analysis is that if some stimulus materials or control conditions have been used more frequently than others, they will have a larger impact on the mean result. Indeed, our literature review shows that some stimulus-materials, like for example false belief vs. photo stories, have been used more often than other stimuli.

A way to address the variability in a meta-analysis is to sort out conceptual or methodological differences between studies, and categorize them along a single stimulus- or task-dimension, as for example verbal versus nonverbal format (Carrington and Bailey,



**Fig. 6.** Regions of interest in fronto-temporal areas. (a) Box-plots show the distributions of effect-sizes; full circles indicate  $p < .005$  uncorrected,  $z > 1$ . (b) Results of pairwise meta-analytic comparisons; full circles indicate  $p < .005$ , Bonferroni corrected. Details same as for Fig. 4.

2009), story-based versus non story-based format (Mar, 2011) or enduring versus transient mental state content (Van Overwalle, 2009). The advantage of this approach is that statistical power is still relatively high because the overall sample is split into just two subsamples. However, focusing on a single stimulus- or content-dimension bears the danger of mixing up effects due to other stimulus- or task-aspects not captured by this one dimension, and these may be of equal importance or interest for the overall results. So, single-dimensioned meta-analyses may again be pooling data across a variety of different stimulus- and task-aspects, which will bring substantial noise to the results. For example, it could be the case that story-based tasks present a certain mental state (e.g., false-belief) more often than non-story based tasks. Or it could be the case that non-story based tasks present a certain form of control condition (e.g., strategic games: participants are told they interact with a computer) more often than story-based tasks. Such

differences would be potential confounds for the comparison between story based and non-story based studies since they have nothing to do with the verbal nature of the tasks. To avoid these problems, it is necessary to separate studies based on the experimental paradigms they used.

#### 4.2. Contribution of the present meta-analysis

For the present meta-analysis, we formed task groups that had comparable stimulus-material, instructions and control conditions. We expected that this approach would reduce the problem of pooling across conceptually and methodologically heterogeneous studies. In support of this assumption, we observed relatively little inter-study heterogeneity in the individual meta-analyses compared to the pooled meta-analysis. No single task-specific meta-analysis showed significant inter-study heterogeneity for

peaks in the right TPJ and the mPFC, and only one or two showed between-study variability for peaks in the left TPJ and the precuneus. A potential drawback of our single task group approach is, however, that producing a number of conceptually homogeneous meta-analyses will inevitably produce smaller samples, which reduces statistical power. We therefore used several strategies to check the quality of our results. We used jack-knife sensitivity analyses to evaluate the robustness of each individual analysis. All meta-analyses were recalculated several times, and each time a different study was excluded from the meta-analytic sample. If results remained unchanged in all recalculations of the meta-analysis, this shows robustness against changes of the sample and thus, indicates that meta-analytic findings are replicable. For example, for our smallest task group (strategic games,  $n=9$ ) we found perfect replicability of findings in left and right TPJ and in the precuneus (9 out of 9 leave-one-out recalculations). In contrast, even in the pooled meta-analysis ( $n=73$ ) not all findings were perfectly replicable, as the left fusiform activation could be only found for 63 out of 73 leave-one-out recalculations.

Most importantly, we took the problem of small sample size into account when testing for overlaps and differences between task groups. For overlaps, we relied on an alternative approach to check for power-problems due to small sample-size. We not only calculated the simple overlap (conjunction) between individual meta-analyses, as this approach may yield a false impression of selectivity. In addition, we carried out a permutation-based overlap analysis, where we determined the overlap between two halves of the pooled-meta analysis sample, each consisting of the studies from three different task groups. This procedure was repeated ten times to create every permutation for assigning six task groups to two sample-halves. Consequently, our procedure identified areas engaged in all task groups, but relied on much larger sample sizes for meta-analyses ( $n > 28$ ). The permutation procedure (see Fig. 1C) turned out to be more sensitive than an analysis of the overlap between the six individual meta-analyses (Fig. 1B). To avoid a false impression of selectivity due to small samples, we will rely on the results from our permutation based procedure in our discussion.

Differences in statistical power must be taken into account when comparing the results of different meta-analyses. For example, when one meta-analysis fails to implicate an area compared to another one, this could reflect differences in sample sizes and between study variability rather than differences between the tasks. To factor this issue in our comparisons, we used a linear model approach which calculates the difference between two meta-analyses while taking into account differences in sample sizes and within- and between study variability. Indeed, our results show a pattern that goes beyond what can be explained by power-differences. For example, when comparing false belief vs. photo tasks with mind in the eyes tasks, we found both areas with overactivation for false belief (e.g., bilateral IPL and TPJp) and areas with overactivation for mind in the eyes (e.g., left IFG). Meta-analytic differences driven by power-differences can be expected only for one direction, i.e., overactivations for the meta-analysis with higher power, but not the other way around.

#### 4.3. Core network for theory of mind

[Mar \(2011\)](#) argued that looking at the overlap between meta-analyses for different paradigm-classes or experiments offers an interesting way for identifying a core-network for theory of mind. The reason is that, whatever is pivotal for theory of mind will be reflected by the convergence of separate methodologies, which differ in idiosyncratic design elements and will not contribute to the overlap. [Mar \(2011\)](#) separated theory of mind research into story based versus non-story based tasks. In the present study, we performed a more detailed separation of task groups. In addition, we

performed a permutation based overlap analysis to ensure that our fine-grained study separation does not lead to a power-problem and thus a false impression of selectivity. Our permutation based analysis found three areas of overlap: The mPFC, left TPJ and right TPJ (connectivity clusters TPJp on both sides). The same results were also found with an alternative statistical procedure for determining overlap between meta-analytic maps, a so-called multimodal meta-analysis ([Radua et al., 2013](#)). The right TPJ cluster showed the most robust overlap among all findings, as it could even be found in a highly conservative and low-powered conjunction analysis. These results support the conclusion from earlier reviews that a ‘core-network’ for theory of mind exists, i.e., that all sorts of theory of mind tasks consistently engage a particular brain network, including mPFC and bilateral TPJ (e.g., [Amodio and Frith, 2006](#); [Frith and Frith, 2006](#); [Mitchell, 2009](#)). This also relates to theories that have proposed a specialized mechanism for mental state attribution, as for example the “theory of mind mechanism” ([ToMM, Leslie and Thaiss, 1992](#)), which was conceived as a mechanism that “kick-starts belief and desire attribution” ([Leslie et al., 2004](#)). Our results, however, suggest that this mechanism also includes attribution of mental states other than beliefs and desires, as we also found activation in mPFC and bilateral TPJ for tasks like mind in the eyes and trait judgments. The question about functional specificity, i.e., the question whether some parts of mPFC and bilateral TPJ are exclusively (or at least preferentially) engaged in theory of mind, goes beyond the scope of our work. In the following sections, we will discuss the functional roles of the core network’s components as well as the implications of task-related differences that we found.

#### 4.4. Areas for theory of mind

##### 4.4.1. Temporo-parietal junction

The TPJ is roughly characterized as an area at the border between the temporal and parietal lobes surrounding the ends of the Sylvian fissure. It is sometimes also referred to as pSTS, posterior IPL, Ventral Parietal Cortex, Angular Gyrus or Brodmann Area 39. This is linked to the lacking consensus on how coordinates, micro- or macroanatomical landmarks should topographically define that area. The functional heterogeneity of this area has become already evident through patient and brain stimulation studies. On the one hand, damage to the right TPJ frequently leads to an inability to attend to stimuli that are presented in the contralateral visual field (i.e., neglect, see [Corbetta and Shulman, 2011](#)) and electrical stimulation of the TPJ results in altered self-awareness. For example, electrical stimulation to the right TPJ leads to the so-called ‘out-of-body experience’ ([Blanke et al., 2002](#)), whereas stimulation to the left TPJ leads to the perception of an ‘illusory shadow person’ ([Arzy et al., 2006](#)). On the other hand, transcranial magnetic stimulation (TMS) of the right TPJ was found to affect belief reasoning: [Young et al. \(2010a\)](#) showed that after TMS, participants were for example impaired in taking into account a person’s intention to harm when asked for a moral evaluation of an act of violence. Moreover, two patient studies ([Apperly et al., 2004](#); [Samson et al., 2004](#)) reported that damage to the left TPJ leads to selective deficits in false belief reasoning, whereas other cognitive and executive domains remained intact.

In the functional imaging field, studies found that temporo-parietal areas contributing to theory of mind can be functionally divided into a more ventral/anterior part including the STS and a more dorsal/posterior part in the TPJ (e.g., [Bahnemann et al., 2010](#); [Gobbini et al., 2007](#); [Perner and Leekam, 2008](#); [Saxe, 2006](#)). However, the only formal literature review looking at task-related differences for theory of mind ([Van Overwalle, 2009](#)) found equal activation in the TPJ across a variety of tasks. A potential explanation for the absence of task-related differences is that [Van Overwalle \(2009\)](#) used one large ROI to analyze the TPJ, which

did not allow checking for differences between anterior versus posterior or dorsal versus ventral parts. The present study used a whole-brain voxel wise meta-analysis that is capable of detecting such differences. Our results show two things: On the one hand, our conjunction analysis found that all theory of mind tasks activate a part of the TPJ which roughly corresponds to area TPJp. On the other hand, we found task-related differences in our ROI analyses (see Fig. 4): More dorsal/posterior areas in the TPJ (TPJp and IPL) showed stronger activation for false belief vs. photo, and tendentially also for trait judgments, compared to other task. More anterior/ventral areas (TPJa and pMTG) showed stronger activations for social animations, mind in the eyes and rational actions compared to other tasks.

The observed functional dissociation within the TPJ supports previous theories (Gobbini et al., 2007; Perner and Leekam, 2008; Perner and Roessler, 2012). For example, Gobbini et al. (2007) linked ventral/anterior TPJ to processing of overt mental states and dorsal/posterior TPJ to processing of covert mental states. Overt mental states are inherent in perceived actions; covert mental states, such as false beliefs, are not necessarily associated with current actions. Perner and Leekam (2008, see also Perner and Roessler, 2010, 2012) explained the functional separation in the TPJ by drawing on the distinction between teleology and belief-desire reasoning which was found in developmental research. Theory of mind tasks are first of all subserved by teleology, we tell another person's intention simply by knowing that a certain action is a rational means to achieve a particular goal. Teleological reasoning (linked to ventral/anterior TPJ) is for example sufficient to understand social animations, mind in the eyes or rational actions. People switch to belief-desire reasoning (linked to dorsal/posterior TPJ) by considering agents' subjective perspectives of what is needed only when required by the task (perspective tasks). In addition to false belief vs. photo, also trait judgments require awareness of perspective, but for different reasons. Traits are habitual patterns of behavior, thought, and emotion. They are characteristic for a person when the person's habits deviate from the norm. For instance, the trait word "stubborn" (Murphy et al., 2010) describes a person who, according to Merriam Webster Dictionary, refuses to change her ideas or to stop doing something. In other words, a stubborn person still sees a point in persisting when objectively (from the judging person's point of view) it is time to give up. In another example, a person is called "anxious" or "nervous" (Mitchell et al., 2002) if she tends to be concerned about situations where one normally has no reason to be anxious, i.e., the person takes a deviant perspective on how dangerous or challenging a situation is. So many traits result from habitually biased perspectives and trait judgments are judgments about whether a person habitually takes a different perspective on certain aspects of life.

The overall pattern found for the TPJ – overlap in the TPJp plus functional differentiations in TPJp and IPL versus TPJa and pMTG – can be linked to a concept of functional specialization referred to as the 'overarching view'. This concept was proposed by Cabeza et al. (2012) with respect to the IPL. The overarching view assumes that although functional subdivisions within a broad brain region exist, they are graded because each subdivision mediates a particular aspect of a global cognitive function supported by the broad region. In our case, the global cognitive function could be inferring/predicting mental states based on (various forms of) available information about a person. Within that broader region (i.e., the TPJ) sub regions apply the global function to different types of information which vary according to patterns of functional connectivity.

We looked at whether the previously found structural connectivity networks of TPJa and TPJp (Mars et al., 2012; Bzdok et al., 2013) are reflected in patterns of task-related co-activations of our ROIs. We labeled our results according to a structural connectivity based parcellation (CBP) atlas by Mars et al. (2011, 2012). This atlas

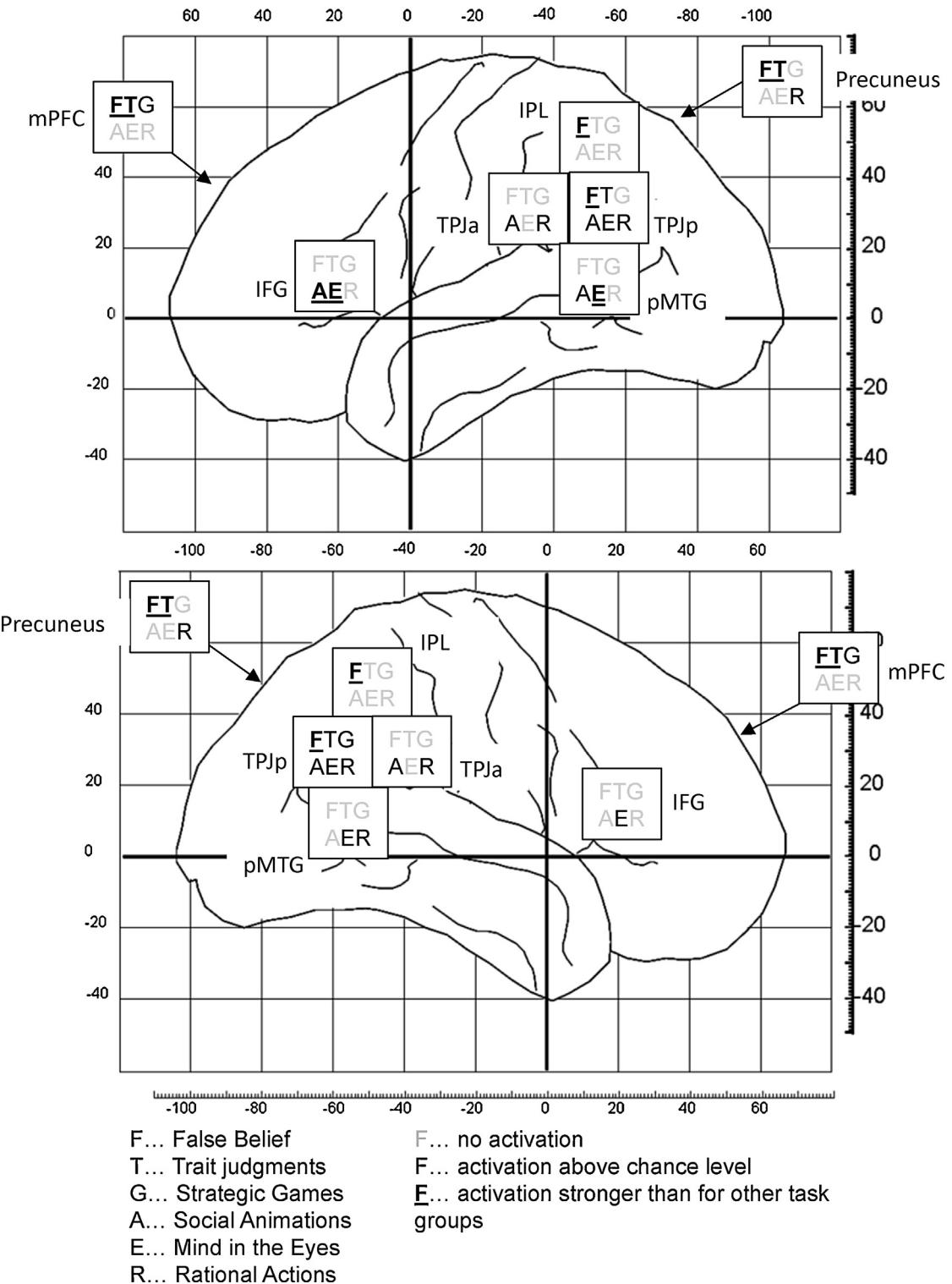
is based on white-matter anatomy. Recently, Bzdok et al. (2013) was able to corroborate this parcellation scheme based on two functional connectivity measures, namely resting-state connectivity analysis and meta-analytic co-activation mapping. Consistently, Mars et al. (2012) and Bzdok et al. (2013) reported that a more anterior portion of the TPJ, labeled as TPJa, is interconnected with inferior frontal gyrus, anterior insula and the SMA. A more posterior portion of the TPJ, the TPJp, is interconnected with dorsal IPL, the precuneus, the ventral mPFC, and posterior middle temporal areas.

When looking at similarities between task-related activation patterns of different ROIs, we focus only on areas that relate to the structural connectivity networks of TPJa (IFG) and TPJp (pMTG, IPL, precuneus and mPFC). As summarized in Fig. 7, activation patterns found in ROIs belonging to the same structural connectivity network do not show a perfect correspondence to each other. However, some broad tendencies are observable. TPJa, IFG and pMTG only activate for some of the task groups: social animations, rational actions and mind in the eyes tasks. As discussed above, these tasks present depictions of human action or behavior. On the other hand, TPJa, IFG and pMTG univocally do not activate for false belief vs. photo, trait judgments, and strategic games. Structural connectivity shows a link between TPJa and IFG. However, the pMTG is linked to TPJp rather than TPJa. The cortical midline areas mPFC and precuneus activate most strongly for false belief vs. photo and trait judgments. The IPL activates only for false belief vs. photo (but tendentially also for trait judgments). The latter three areas are all structurally linked to TPJp, which also activates strongly for false belief vs. photo and at an average level for trait judgments. However, the TPJp does not only activate for the latter two task groups, but also for social animations, mind in the eyes and rational actions, which could be due to its connections to the pMTG. Taken together, our ROIs show some support for the idea that the nature and location of activation in the TPJ is, at least in part, reflected in the network of co-activated areas. Original imaging studies are necessary to further clarify the extent to which activation in TPJa versus TPJp can be predicted by connectivity-networks.

#### 4.4.2. Medial prefrontal cortex

Strong activation of the mPFC has already been recognized in early studies on theory of mind, which led to the initial assumption that it is specifically linked to reasoning about belief (Frith and Frith, 2001; Gallagher and Frith, 2003; Leslie et al., 2004). However, this hypothesis has been put into question by more recent patient and imaging data. Bird et al. (2004) reported a case of a patient with an extensive medial prefrontal lesion but no impairment across a wide range of theory of mind tasks, including the 'strange stories' test which probes false-belief understanding (see Jolliffe and Baron-Cohen, 1999). Furthermore, while it was found that damage to the left TPJ leads to selective deficits in reasoning about belief, medial frontal damage leads to deficits both in belief reasoning and other cognitive and executive tasks (Apperly et al., 2004; Samson et al., 2004). In a functional imaging study, it was found that the mPFC is equally engaged by stories about persons' thoughts and by stories about physical appearance or bodily sensations (Saxe and Powell, 2006). Based on this evidence Saxe and Powell (2006) suggested that the mPFC is not specifically engaged in belief-desire reasoning, but more generally involved in processing socially or emotionally relevant information about others (see also Aichhorn et al., 2006). Behrens et al. (2008) showed that the mPFC uses information about others to compute an overall impression of another's personality. During interactive games, Behrens et al. (2008) found that trial-by-trial changes in activation in the mPFC reflect learning about the 'social value' of another player, e.g., the gradual build-up of a prediction about how reliable the other player is. This is consistent with the idea that mPFC is particularly engaged when participants are asked to make judgments about a person's enduring psychological

## Schematic summary of main results from ROI-based analysis



**Fig. 7.** Schematic summary of significant activations in ROIs that are linked to the structural connectivity networks of TPJa and TPJp. Labels for ROIs are the same as in previous figures. Task groups are abbreviated by capital letters. Black font indicates that mean-analysis shows significant activation for a task group in that ROI ( $p < .005$  uncorrected,  $z > 1$ ). Underlined letters indicate that linear-contrasts found stronger activation for this task group compared to others in an area ( $p < .005$ , Bonferroni corrected for the number of pairwise comparisons).

and social properties, such as for example personality traits (Van Overwalle, 2009). Interestingly, Krienen et al. (2010) has recently shown that mPFC activation during trait judgments is driven by the social closeness of the target-person (e.g., friend or stranger?) rather than by mental or psychological characteristics (e.g., how similar are these traits to my own?).

Although our permutation based conjunction showed that dorsal parts of the mPFC are engaged by all theory of mind tasks, we found in our ROI analyses (see Fig. 5) that activation is particularly strong for false belief vs. photo, strategic games and trait judgment tasks, and less reliable for social animations, mind in the eyes and rational actions. The relevant difference between these tasks may be the time-scale on which socially relevant information is processed. False-belief stories often tell about multiple events in a person's life, and strategic games present the same opponent across multiple rounds of the game. Trait judgments explicitly ask to think about enduring personality traits. On the other hand, social animations, mind in the eyes and rational actions often show only a single action, behavior or facial expression. Presumably, such information is less relevant for reasoning about a person's stable personality traits or her 'social value'.

#### 4.4.3. Precuneus

Little evidence concerning the precuneus exists from patient and brain stimulation research, mainly because of its hidden location within the parietal lobe. Despite strong agreement on the fact that activation of the precuneus is one of the most robust correlates of theory of mind, the area has received relatively little attention in functional accounts. Convergent research shows that the precuneus is implicated in visuo-spatial mental imagery. The precuneus is engaged by motor imagery (e.g., Hanakawa et al., 2003), by mental rotation (Suchan et al., 2002), mental simulation of routes (Ghaem et al., 1997), and by mental model building in deductive reasoning tasks (e.g., Knauff et al., 2003; Kulakova et al., 2013). In addition, the precuneus is engaged in judgments about another person's visual perspective (Vogeley et al., 2004). Based on these findings, Cavanna and Trimble (2006) have proposed that one main function of the precuneus in theory of mind is mental imagery to represent the perspective of another person. Results from our ROI analysis (see Fig. 5) show strong activation for false belief vs. photo, trait-judgments and rational actions. These tasks can all be linked to mental imagery. False belief vs. photo and rational action tasks often tell a fictional story or show a fictional event, which invites mental imaging. Traits are, to some extent, reflected in the habitual patterns of behavior of a person. Therefore, making trait judgments also induces mental imagery of how a person acts or behaves in fictitious situations. On the other hand, social animations and mind in the eyes most clearly do not activate the precuneus, which is consistent with a mental imagery function, because for the latter two tasks the relevant action is shown and does not need to be imagined.

#### 4.4.4. Anterior temporal lobes

Studies of semantic dementia patients (for a review, see Patterson et al., 2007) and transcranial magnetic stimulation (e.g., Lambon Ralph et al., 2009) showed that the anterior temporal lobes are critical for the storage of semantic knowledge. For example, it is assumed that the anterior temporal lobes store semantic concepts which allow generalization of knowledge from one exemplar to another (Lambon Ralph and Patterson, 2008). With respect to social cognition, researchers suggested that the anterior temporal lobes are important for storage and retrieval of social semantic scripts (e.g., Frith and Frith, 2003; Gallagher and Frith, 2003). A script refers to general world knowledge which can be seen as a record of which particular activities will take place in a particular setting. An example was given by Gallagher and Frith (2003, p. 77): '... the activities associated with the restaurant script would

include reading the menu, ordering a drink and getting the bill. If I catch the waiter's eye and make the gesture of writing on my left palm with my right forefinger he will usually bring me the bill. He correctly interprets my actions on the basis of his knowledge of my likely goals...'.

Recent research has extended this line of reasoning, and arrived at the conclusion that the anterior temporal lobes provide not only scripts but social semantic concepts in general. For example, Zahn et al. (2007) asked participants to judge the semantic similarity of word pairs which were either descriptive of human social semantic concepts ('honor' – 'brave') or describing biological function ('nutritious' – 'useful'). Social semantic concepts were specifically associated with an increase in activation in bilateral anterior temporal lobes. Sugiura et al. (2006) showed that the anterior temporal lobes are also important areas for recognition of personally familiar or known persons, which is based on relating semantic information to persons (see also Tsukiura et al., 2010). Ross and Olson (2010) showed that overlapping areas of the anterior temporal lobes are engaged by social animations (comparable to those in our meta-analysis), semantic judgments on social words (comparable to those presented during trait judgments in our meta-analysis), and by stories about persons which require reasoning about intentions. The authors concluded that the anterior temporal lobes contribute to the understanding of implied meaning through access to both general conceptual knowledge and to specific social conceptual knowledge, such as background knowledge about social descriptors (e.g., words like friendly and devious), and knowledge about social rules and social etiquette. Findings from our meta-analysis (see Fig. 6) are in support of this social semantic concept interpretation. Activation in the anterior temporal lobes was found for trait judgments containing social concept words, as well as for false belief vs. photo social animations and rational actions, which all contain a sequence of actions in a social script (e.g., Gallagher and Frith, 2003; Olson et al., 2007). In addition, some of these task groups also activated more central areas in the middle temporal lobe, whereas activation for trait judgments was restricted to the anterior temporal lobes.

#### 4.4.5. Inferior frontal gyrus

The present meta-analysis found convergent activation in the left inferior frontal gyrus only for the task groups social animations and mind in the eyes. For both types of tasks, evidence has suggested that inferior frontal mirror neuron circuits aid in identification of the observed actions and emotional expressions by a 'common coding' mechanism for action and perception (see e.g., Keysers et al., 2010; Keysers and Gazzola, 2007). Functional imaging research shows that the inferior frontal gyrus is engaged in both the execution of actions and in the observation of these actions performed by someone else (for review, see Rizzolatti and Craighero, 2004). Furthermore, Carr et al. (2003) showed that these inferior frontal areas support emotion identification and empathy due to their connections to the limbic system.

The functional significance of the inferior frontal gyrus for processing of abstract displays of implied biological movements was shown in a study by Saygin (2007). The authors found that a group of unilateral stroke patients with lesions to either the left or the right inferior frontal gyrus were severely impaired in correctly identifying the human actions depicted by point-light displays (e.g., walking, jogging, ...). With functional imaging, the authors additionally showed that these frontal areas were activated in nonimpaired adults for the identification of actions depicted by point-light displays. With respect to the mind in the eyes task, Shamay-Tsoory et al. (2009) found that a group of patients with lesions in the inferior frontal gyrus had particular difficulties here, but performed unimpaired on a second-order false belief task (Stone et al., 1998). In addition, the patients scored relatively low

on affective empathy scales of the Interpersonal Reactivity Index Questionnaire (Davis, 1983), namely the scales 'empathic concern' (e.g., 'I often have tender, concerned feelings for people less fortunate than me') and 'personal distress' (e.g., 'being in a tense emotional situation scares me'). Taken together, the patient studies highlight that the inferior frontal gyrus is a crucial structure for particular forms of theory of mind reasoning, which two out of the six task groups from our meta-analysis tap into.

#### 4.5. Limitations

This meta-analysis is intended to give an overview of the rapidly growing field of imaging research on theory of mind. Our focus on standard task-based neuroimaging data dismisses other evidence which is crucial for understanding the neural underpinnings of theory of mind. First of all, the reviewed imaging data, essentially correlational in nature, should always be supplemented by data related to causality, such as TMS and brain lesion studies. We have incorporated both kinds of evidence in our discussion of brain areas to highlight the importance of causal considerations. Second, our meta-analytic procedure excluded findings from advanced methods of analysis. For example, some studies with strategic games used computational modeling for data analysis (Yoshida et al., 2010; Behrens et al., 2008; Hampton et al., 2008). These studies found that the pSTS computes discrepancies between the behavior that one expects from another player (indicated by the nature of one's own decision) and her actual behavior (Behrens et al., 2008; Hampton et al., 2008). The mPFC, on the other hand, computes (i.e., learning) the social "value" of another player, e.g., how much one can trust that person (Behrens et al., 2008).

### 5. Conclusion

For the present meta-analysis, we formed task groups that had comparable stimulus-material, instructions and control conditions. Meta-analytic assessments of between-study heterogeneity showed that our approach reduces the problem of pooling across conceptually and methodologically different studies. We carried out overlap analyses between task groups, and found that the mPFC and bilateral posterior TPJ (connectivity cluster TPJp) showed activation for all theory of mind tasks. This is in line with claims about the existence of a 'core-network' for theory of mind, i.e., that all sorts of theory of mind tasks consistently engage a particular brain network (e.g., Amodio and Frith, 2006; Frith and Frith, 2006; Mitchell, 2009). In addition, we performed ROI analysis and found a number of task-related activation differences along with the core-network. For example, we found that more dorsal/posterior parts of the TPJ are particularly engaged for tasks that require processing of mental perspectives (Perner and Leekam, 2008; Perner and Roessler, 2012), namely false belief vs. photo and tendentially trait judgments. More ventral/anterior parts of the TPJ are preferentially activated by tasks which depict rational actions or behavior. Together, this pattern of overlap and surrounding activation differences can be linked to a concept of functional specialization referred to as the 'overarching view' (Cabeza et al., 2012), which has been suggested for the IPL. This view assumes that although functional subdivisions within a broad brain region exist, they are graded because each subdivision mediates a particular aspect of a global cognitive function supported by the broad region. In our case, the global cognitive function could be inferring/predicting unobservable mental states based on (various forms of) available information about a person. Sub regions apply this global function to different types of information which vary according to patterns of functional and structural brain connectivity. In support of the connectivity-view, we found that similarities between ROI-activation patterns

are consistent patterns with structural brain connectivity networks of TPJ subdivisions (see Mars et al., 2012).

### Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.neubiorev.2014.01.009>.

### References

- Aichhorn, M., Perner, J., Kronbichler, M., Staffen, W., Ladurner, G., 2006. Do visual perspective tasks need theory of mind? *Neuroimage* 30, 1059–1068.
- Amodio, D.M., Frith, C.D., 2006. Meeting of minds: the medial frontal cortex and social cognition. *Nature Review Neuroscience* 7, 268–277.
- Apperly, I.A., Samson, D., Chiavarino, C., Humphreys, G.W., 2004. Frontal and temporo-parietal lobe contributions to theory of mind: neuropsychological evidence from a false-belief task with reduced language and executive demands. *Journal of Cognitive Neuroscience* 16, 1773–1784.
- Arzy, S., Seck, M., Ortigue, S., Spinelli, L., Blanke, O., 2006. Induction of an illusory shadow person. *Nature* 443, 287.
- Baglio, F., Castelli, I., Alberoni, M., Blasi, V., Griffanti, L., Falini, A., Nemni, R., Marchetti, A., 2012. Theory of mind in amnestic mild cognitive impairment: an fMRI study. *Journal of Alzheimer's Disease* 29, 25–37.
- Bahnemann, M., Dziobek, I., Prehn, K., Wolf, I., Heekerlen, H.R., 2010. Sociotopy in the temporoparietal cortex: common versus distinct processes. *Social Cognitive and Affective Neuroscience* 5, 48–58.
- Baron-Cohen, S., Hammer, J., 1997. Parents of children with Asperger syndrome: what is the cognitive phenotype? *Journal of Cognitive Neuroscience* 9, 548–554.
- Baron-Cohen, S., Jolliffe, T., Mortimore, C., Robertson, M., 1997. Another advanced test of theory of mind: evidence from very high functioning adults with autism or Asperger syndrome. *Journal of Child Psychology and Psychiatry* 38, 813–822.
- Baron-Cohen, S., Ring, H.A., Moriarty, J., Schmitz, B., Costa, D., Ell, P., 1994. Recognition of mental state terms. Clinical findings in children with autism and a functional neuroimaging study of normal adults. *British Journal of Psychiatry* 165, 640–649.
- Baron-Cohen, S., Ring, H.A., Wheelwright, S., Bullmore, E.T., Brammer, M.J., Simmons, A., Williams, S.C., 1999. Social intelligence in the normal and autistic brain: an fMRI study. *European Journal of Neuroscience* 11, 1891–1898.
- Behrens, T.E., Hunt, L.T., Woolrich, M.W., Rushworth, M.F., 2008. Associative learning of social value. *Nature* 456, 245–249, <http://dx.doi.org/10.1038/nature07538>.
- Bhatt, M., Camerer, C.F., 2005. Self-referential thinking and equilibrium as states of mind in games: fMRI evidence. *Games and Economic Behavior* 81, 145–164, <http://dx.doi.org/10.1016/j.geb.2005.03.007>.
- Blackwood, N.J., Bentall, R.P., ffytche, D.H., Simmons, A., Murray, R.M., Howard, R.J., 2003. Self-responsibility and the self-serving bias: an fMRI investigation of causal attributions. *Neuroimage* 20, 1076–1085.
- Blakemore, S.J., Fonlupt, P., Pachot-Clouard, M., Darmon, C., Boyer, P., Meltzoff, A.N., Segebarth, C., Decety, J., 2001. How the brain perceives causality: an event-related fMRI study. *Neuroreport* 12, 3741–3746.
- Blanke, O., Ortigue, S., Landis, T., Seck, M., 2002. Stimulating illusory own-body perceptions. *Nature* 419, 269–270.
- Bird, C.M., Castelli, F., Malik, O., Frith, U., Husain, M., 2004. The impact of extensive medial frontal lobe damage on "Theory of Mind" and cognition. *Brain* 127, 914–928.
- Brunet, E., Sarfati, Y., Hardy-Baylé, M.C., 2003. Reasoning about physical causality and other's intentions in schizophrenia. *Cognitive Neuropsychiatry* 8, 129–139.
- Brunet, E., Sarfati, Y., Hardy-Baylé, M.C., Decety, J., 2000. A PET investigation of the attribution of intentions with a nonverbal task. *Neuroimage* 11, 157–166.
- Brüne, M., Ozgürdal, S., Ansorge, N., von Reventlow, H.G., Peters, S., Nicolas, V., Tegenthoff, M., Juckel, G., Lissek, S., 2011. An fMRI study of "theory of mind" in at-risk states of psychosis: comparison with manifest schizophrenia and healthy controls. *Neuroimage* 55, 329–337.
- Bzdok, D., Langner, R., Schilbach, L., Jakobs, O., Roski, C., Caspers, S., Laird, A.R., Fox, P.T., Zilles, K., Eickhoff, S.B., 2013. Characterization of the temporo-parietal junction by combining data-driven parcellation, complementary connectivity analyses, and functional decoding. *Neuroimage* 81, 381–392.
- Bzdok, D., Schilbach, L., Vogeley, K., Schneider, K., Laird, A.R., Langner, R., Eickhoff, S.B., 2012. Parsing the neural correlates of moral cognition: ALE meta-analysis on morality, theory of mind, and empathy. *Brain Structure and Function* 217, 783–796.
- Cabeza, R., Ciaramelli, E., Moscovitch, M., 2012. Cognitive contributions of the ventral parietal cortex: an integrative theoretical account. *Trends in Cognitive Sciences* 16, 338–352, <http://dx.doi.org/10.1016/j.tics.2012.04.008>.
- Carr, L., Iacoboni, M., Dubeau, M.C., Mazziotta, J.C., Lenzi, G.L., 2003. Neural mechanisms of empathy in humans: a relay from neural systems for imitation to limbic areas. *Proceedings of the National Academy of Sciences of the United States of America* 100, 5497–5502.
- Carrington, S.J., Bailey, A.J., 2009. Are there theory of mind regions in the brain? A review of the neuroimaging literature. *Human Brain Mapping* 30, 2313–2335.

- Castelli, F., Happé, F., Frith, U., Frith, C.D., 2000. Movement and mind: a functional imaging study of perception and interpretation of complex intentional movement patterns. *Neuroimage* 12, 314–325.
- Castelli, F., Frith, C.D., Happé, F., Frith, U., 2002. Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes. *Brain* 125, 1839–1849.
- Cavanna, A.E., Trimble, M.R., 2006. The precuneus: a review of its functional anatomy and behavioural correlates. *Brain* 129, 564–583.
- Ciaramidaro, A., Adenzato, M., Enrici, I., Erk, S., Pia, L., Bara, B.G., Walter, H., 2007. The intentional network: how the brain reads varieties of intentions. *Neuropsychologia* 45, 3105–3113.
- Cloutier, J., Kelley, W.M., Heatherton, T.F., 2011. The influence of perceptual and knowledge-based familiarity on the neural substrates of face perception. *Social Neuroscience* 6, 63–75, <http://dx.doi.org/10.1080/17470911003693622>.
- Cooper, D., Barker, V., Radua, J., Fusar-Poli, P., Lawrie, S., 2013. Multimodal voxel-based meta-analysis of structural and functional magnetic resonance imaging studies in those at elevated genetic risk of developing schizophrenia. *Psychiatry Research* 221, 69–77, <http://dx.doi.org/10.1016/j.psychresns.2013.07.008>.
- Corbetta, M., Shulman, G.L., 2011. Spatial neglect and attention networks. *Annual Review of Neuroscience* 34, 569–599.
- D'Argembeau, A., Ruby, P., Collette, F., Degueldre, C., Balteau, E., Luxen, A., Maquet, P., Salmon, E., 2007. Distinct regions of the medial prefrontal cortex are associated with self-referential processing and perspective taking. *Journal of Cognitive Neuroscience* 19, 935–944.
- Davis, M.H., 1983. Measuring individual differences in empathy: evidence for a multidimensional approach. *Journal of Personality and Social Psychology* 44, 113–126.
- David, S.P., Ware, J.J., Chu, I.M., Loftus, P.D., Fusar-Poli, P., Radua, J., Munafó, M.R., Ioannidis, J.P.A., 2013. Potential Reporting Bias in fMRI Studies of the Brain. *PLoS One* 8, e70104.
- Decety, J., Jackson, P.L., Sommerville, J.A., Chaminade, T., Meltzoff, A.N., 2004. The neural bases of cooperation and competition: an fMRI investigation. *Neuroimage* 23, 744–751.
- Decety, J., Lamm, C., 2007. The role of the right temporoparietal junction in social interaction: how low-level computational processes contribute to meta-cognition. *Neuroscientist* 13, 580–593.
- Denny, B.T., Kober, H., Wager, T.D., Ochsner, K.N., 2012. A meta-analysis of functional neuroimaging studies of self- and other judgments reveals a spatial gradient for mentalizing in medial prefrontal cortex. *Journal of Cognitive Neuroscience* 24, 1742–1752, <http://dx.doi.org/10.1162/jocn.a.00233>.
- Dennett, D.C., 1971. Intentional systems. *Journal of Philosophy* 8, 87–106.
- DerSimonian, R., Laird, N., 1986. Meta-analysis in clinical trials. *Controlled Clinical Trials* 7, 177–188.
- Doehnel, K., Schuwerk, T., Meinhardt, J., Sodian, B., Hajak, G., Sommer, M., 2012. Functional activity of the right temporo-parietal junction and of the medial prefrontal cortex associated with true and false belief reasoning. *Neuroimage* 60, 1652–1661, <http://dx.doi.org/10.1016/j.neuroimage.2012.01.073>.
- Eickhoff, S.B., Laird, A.R., Grefkes, C., Wang, L.E., Zilles, K., Fox, P.T., 2009. Coordinate-based activation likelihood estimation meta-analysis of neuroimaging data: a random-effects approach based on empirical estimates of spatial uncertainty. *Human Brain Mapping* 30, 2907–2926, <http://dx.doi.org/10.1002/hbm.20718>.
- Elliott, R., Völlm, B., Drury, A., McKie, S., Richardson, P., Deakin, J.F., 2006. Co-operation with another player in a financially rewarded guessing game activates regions implicated in theory of mind. *Social Neuroscience* 1, 385–395, <http://dx.doi.org/10.1080/17470910601041358>.
- Fletcher, P.C., Happé, F., Frith, U., Baker, S.C., Dolan, R.J., Frackowiak, R.S., Frith, C.D., 1995. Other minds in the brain: a functional imaging study of "theory of mind" in story comprehension. *Cognition* 57, 109–128.
- Fonlupt, P., 2003. Perception and judgement of physical causality involve different brain structures. *Cognitive Brain Research* 17, 248–254.
- Frith, C.D., Frith, U., 1999. Interacting minds—a biological basis. *Science* 286, 1692–1695.
- Frith, U., Frith, C.D., 2001. The biological basis of social interaction. *Current Directions in Psychological Science* 10, 151–155.
- Frith, C.D., Frith, U., 2006. The neural basis of mentalizing. *Neuron* 50, 531–534.
- Frith, U., Frith, C.D., 2003. Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society B: Biological Sciences* 358, 459–473.
- Gallagher, H.L., Happé, 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, 11–21.
- Gallagher, H.L., Frith, C.D., 2003. Functional imaging of 'theory of mind'. *Trends in Cognitive Sciences* 7, 77–83.
- Gallagher, H.L., Jack, A.I., Roepstorff, A., Frith, C.D., 2002. Imaging the intentional stance in a competitive game. *Neuroimage* 16, 814–821.
- German, T.P., Niehaus, J.L., Roarty, M.P., Giesbrecht, B., Miller, M.B., 2004. Neural correlates of detecting pretense: automatic engagement of the intentional stance under covert conditions. *Journal of Cognitive Neuroscience* 16, 1805–1817.
- Ghaem, O., Mellet, E., Crivello, F., Tzourio, N., Mazoyer, B., Berthoz, A., Denis, M., 1997. Mental navigation along memorized routes activates the hippocampus, precuneus, and insula. *Neuroreport* 8, 739–744.
- Gobbini, M.I., Koralek, A.C., Bryan, R.E., Montgomery, K.J., Haxby, J.V., 2007. Two takes on the social brain: a comparison of theory of mind tasks. *Journal of Cognitive Neuroscience* 19, 1803–1814.
- Grézes, J., Berthoz, S., Passingham, R.E., 2006. Amygdala activation when one is the target of deceit: did he lie to you or to someone else? *Neuroimage* 30, 601–608.
- Grézes, J., Frith, C.D., Passingham, R.E., 2004a. Brain mechanisms for inferring deceit in the actions of others. *Journal of Neuroscience* 24, 5500–5505.
- Grézes, J., Frith, C.D., Passingham, R.E., 2004b. Inferring false beliefs from the actions of oneself and others: an fMRI study. *Neuroimage* 21, 744–750.
- Gutches, A.H., Kensinger, E.A., Schacter, D.L., 2007. Aging, self-referencing, and medial prefrontal cortex. *Social Neuroscience* 2, 117–133, <http://dx.doi.org/10.1080/17470910701399029>.
- Hampton, A.N., Bossaerts, P., O'Doherty, J.P., 2008. Neural correlates of mentalizing-related computations during strategic interactions in humans. *Proceedings of the National Academy of Sciences of the United States of America* 105, 6741–6746, <http://dx.doi.org/10.1073/pnas.0711099105>.
- Hanakawa, T., Immisch, I., Toma, K., Dimyan, M.A., Van Gelderen, P., Hallett, M., 2003. Functional properties of brain areas associated with motor execution and imagery. *Journal of Neurophysiology* 89, 989–1002.
- Happé, F., Ehlers, S., Fletcher, P., Frith, U., Johansson, M., Gillberg, C., Dolan, R., Frackowiak, R., Frith, C.D., 1996. Theory of mind' in the brain. Evidence from a PET scan study of Asperger syndrome. *NeuroReport* 8, 197–201.
- Heberlein, A.S., Saxe, R.R., 2005. Dissociation between emotion and personality judgments: convergent evidence from functional neuroimaging. *Neuroimage* 28, 770–777.
- Heider, F., Simmel, M., 1944. An experimental study of apparent behavior. *American Journal of Psychology* 57, 243–259.
- Hooker, C.I., Verosky, S.C., Germine, L.T., Knight, R.T., D'Esposito, M., 2008. Mentalizing about emotion and its relationship to empathy. *Social Cognitive and Affective Neuroscience* 3, 204–217, <http://dx.doi.org/10.1093/scan/hsn019>.
- Jenkins, A.C., Mitchell, J.P., 2010. Mentalizing under uncertainty: dissociated neural responses to ambiguous and unambiguous mental state inferences. *Cerebral Cortex* 20, 404–410, <http://dx.doi.org/10.1093/cercor/bhp109>.
- Jolliffe, T., Baron-Cohen, S., 1999. The Strange Stories Test: a replication with high-functioning adults with autism or Asperger syndrome. *Journal of Autism and Developmental Disorders* 29, 395–406.
- Keyser, C., Gazzola, V., 2007. Integrating simulation and theory of mind: from self to social cognition. *Trends in Cognitive Sciences* 11, 194–196.
- Keyser, C., Kaas, J., Gazzola, V., 2010. Somatosensation in social cognition. *Nature Reviews Neuroscience* 11, 417–428.
- Knauff, M., Fangmeier, T., Ruff, C.C., Johnson-Laird, P.N., 2003. Reasoning, models, and images: behavioral measures and cortical activity. *Journal of Cognitive Neuroscience* 15, 559–573.
- Krienen, F.M., Tu, P.C., Buckner, R.L., 2010. Clan mentality: evidence that the medial prefrontal cortex responds to close others. *Journal of Neuroscience* 30, 13906–13915.
- Krill, A.L., Platek, S.M., 2012. Working together may be better: activation of reward centers during a cooperative maze task. *PLoS One* 7, e30613, <http://dx.doi.org/10.1371/journal.pone.0030613>.
- Kulakova, E., Aichhorn, M., Schurz, M., Kronbichler, M., Perner, J., 2013. Processing counterfactual and hypothetical conditionals: an fMRI investigation. *Neuroimage* 15, 265–271.
- Lambon Ralph, M.A., Patterson, K., 2008. Generalization and differentiation in semantic memory: insights from semantic dementia. *Annals of the New York Academy of Sciences* 1124, 61–76, <http://dx.doi.org/10.1196/annals.1440.006>.
- Lambon Ralph, M.A., Pobric, G., Jeffries, E., 2009. Conceptual knowledge is underpinned by temporal pole bilaterally: Convergent evidence from rTMS. *Cerebral Cortex* 19, 832–838, <http://dx.doi.org/10.1093/cercor/bhn131>.
- Lancaster, J.L., Tordesillas-Gutiérrez, D., Martínez, M., Salinas, F., Evans, A., Zilles, K., Mazzotti, J.C., Fox, P.T., 2007. Bias between MNI and Talairach coordinates analyzed using the ICBM-152 brain template. *Human Brain Mapping* 28, 1194–1205.
- Leslie, A.M., Friedman, O., German, T.P., 2004. Core mechanisms in 'theory of mind'. *Trends in Cognitive Sciences* 8, 528–533.
- Leslie, A.M., Thaiss, L., 1992. Domain specificity in conceptual development: neuropsychological evidence from autism. *Cognition* 43, 225–251.
- Lombardo, M.V., Chakrabarti, B., Bullmore, E.T., Wheelwright, S.J., Sadek, S.A., Suckling, J., MRC AIMS Consortium, Baron-Cohen, S., 2010. Shared neural circuits for mentalizing about the self and others. *Journal of Cognitive Neuroscience* 22, 1623–1635, <http://dx.doi.org/10.1162/jocn.2009.21287>.
- Mar, R.A., 2011. The neural bases of social cognition and story comprehension. *Annual Review of Psychology* 62, 103–134, <http://dx.doi.org/10.1146/annurev-psych-120709-145406>.
- Mars, R.B., Jbabdi, S., Sallet, J., O'Reilly, J.X., Croxson, P.L., Olivier, E., Noonan, M.P., Bergmann, C., Mitchell, A.S., Baxter, M.G., Behrens, T.E., Johansen-Berg, H., Tomassini, V., Miller, K.L., Rushworth, M.F., 2011. Diffusion-weighted imaging tractography-based parcellation of the human parietal cortex and comparison with human and macaque resting-state functional connectivity. *Journal of Neuroscience* 31, 4087–4100, <http://dx.doi.org/10.1523/JNEUROSCI.5102-10.2011>.
- Mars, R.B., Sallet, J., Schüffelgen, U., Jbabdi, S., Toni, I., Rushworth, M.F., 2012. Connectivity-based subdivisions of the human right "temporoparietal junction area": evidence for different areas participating in different cortical networks. *Cerebral Cortex* 22, 1894–1903.
- Marjoram, D., Job, D.E., Whalley, H.C., Gountouna, V.E., McIntosh, A.M., Simonotto, E., Cunningham-Owens, D., Johnstone, E.C., Lawrie, S., 2006. A visual joke fMRI investigation into theory of mind and enhanced risk of schizophrenia. *Neuroimage* 31, 1850–1858.
- Mier, D., Lis, S., Neuthe, K., Sauer, C., Esslinger, C., Gallhofer, B., Kirsch, P., 2010. The involvement of emotion recognition in affective theory of mind. *Psychophysiology* 47, 1028–1039, <http://dx.doi.org/10.1111/j.1469-8986.2010.01031.x>.

- Mitchell, J.P., 2009. Inferences about mental states. *Philosophical Transactions of the Royal Society B: Biological Sciences* 364, 1309–1316, <http://dx.doi.org/10.1098/rstb.2008.0318>.
- Mitchell, J.P., Heatherton, T.F., Macrae, C.N., 2002. Distinct neural systems subserve person and object knowledge. *Proceedings of the National Academy of Sciences of the United States of America* 99, 15238–15243.
- Modinos, G., Ormel, J., Aleman, A., 2009. Activation of anterior insula during self-reflection. *PLoS One* 4, e4618, <http://dx.doi.org/10.1371/journal.pone.0004618>.
- Murphy, E.R., Brent, B.K., Benton, M., Pruitt, P., Diwadkar, V., Rajarethnam, R.P., Keshavan, M.S., 2010. Differential processing of metacognitive evaluation and the neural circuitry of the self and others in schizophrenia: a pilot study. *Schizophrenia Research* 116, 252–258, <http://dx.doi.org/10.1016/j.schres.2009.11.009>.
- Murray, R.J., Schaer, M., Debbané, M., 2012. Degrees of separation: a quantitative neuroimaging meta-analysis investigating self-specificity and shared neural activation between self- and other-reflection. *Neuroscience & Biobehavioral Reviews* 36, 1043–1059, <http://dx.doi.org/10.1016/j.neubiorev.2011.12.013>.
- Nichols, T., Brett, M., Andersson, J., Wager, T., Poline, J.B., 2005. Valid conjunction inference with the minimum statistic. *Neuroimage* 25, 653–660, <http://dx.doi.org/10.1016/j.neuroimage.2004.12.005>.
- Olson, I.R., Plotzker, A., Ezzyat, Y., 2007. The enigmatic temporal pole: a review of findings on social and emotional processing. *Brain* 130, 1718–1731.
- Patterson, C., Feightner, J., Garcia, A., MacKnight, K., 2007. General risk factors for dementia: a systematic evidence review. *Alzheimer's & Dementia* 3, 341–347, <http://dx.doi.org/10.1016/j.jalz.2007.07.001>.
- Perner, J., Leekam, S., 2008. The curious incident of the photo that was accused of being false: issues of domain specificity in development, autism, and brain imaging. *Quarterly Journal of Experimental Psychology* 61, 76–89.
- Perner, J., Roessler, J., 2010. Teleology and causal reasoning in children's theory of mind. In: Aguilar, J., Buckareff, A.A. (Eds.), *Causing Human Action: New Perspectives on the Causal Theory of Action*. Bradford Book, The MIT Press, Cambridge, MA, pp. 199–228.
- Perner, J., Roessler, J., 2012. From infants' to children's appreciation of belief. *Trends in Cognitive Sciences* 16, 519–525, <http://dx.doi.org/10.1016/j.tics.2012.08.004>.
- Péron, J., Le Jeune, F., Haegelen, C., Dondaine, T., Drapier, D., Sauleau, P., Reymann, J.M., Drapier, S., Rouaud, T., Millet, B., Vérin, M., 2010. Subthalamic nucleus stimulation affects theory of mind network: a PET study in Parkinson's disease. *PLoS One* 5, e9919, <http://dx.doi.org/10.1371/journal.pone.0009919>.
- Pfeifer, J.H., Lieberman, M.D., Dapretto, M., 2007. "I know you are but what am I!": neural bases of self- and social knowledge retrieval in children and adults. *Journal of Cognitive Neuroscience* 19, 1323–1337.
- Radua, J., Mataix-Cols, D., Phillips, M.L., El-Hage, W., Kronhaus, D.M., Cardoner, N., Surguladze, S., 2012. A new meta-analytic method for neuroimaging studies that combines reported peak coordinates and statistical parametric maps. *European Psychiatry* 27, 605–611, <http://dx.doi.org/10.1016/j.eurpsy.2011.04.001>.
- Radua, J., Romeo, M., Mataix-Cols, D., Fusar-Poli, P., 2013. A general approach for combining voxel-based meta-analyses conducted in different neuroimaging modalities. *Current Medicinal Chemistry* 20, 462–466.
- Ray, R.D., Shelton, A.L., Hollon, N.G., Matsumoto, D., Frankel, C.B., Gross, J.J., Gabrieli, J.D., 2010. Interdependent self-construal and neural representations of self and mother. *Social Cognitive and Affective Neuroscience* 5, 318–323, <http://dx.doi.org/10.1093/scan/nsp039>.
- Rilling, J.K., Goldsmith, D.R., Glenn, A.L., Jairam, M.R., Elfenbein, H.A., Dagenais, J.E., Murdock, C.D., Pagnoni, G., 2008. The neural correlates of the affective response to unreciprocated cooperation. *Neuropsychologia* 46, 1256–1266, <http://dx.doi.org/10.1016/j.neuropsychologia.2007.11.033>.
- Rilling, J.K., Sanfey, A.G., Aronson, J.A., Nystrom, L.E., Cohen, J.D., 2004. The neural correlates of theory of mind within interpersonal interactions. *Neuroimage* 22, 1694–1703.
- Rizzolatti, G., Craighero, L., 2004. The mirror-neuron system. *Annual Review of Neuroscience* 27, 169–192.
- Ross, L.A., Olson, I.R., 2010. Social cognition and the anterior temporal lobes. *Neuroimage* 49, 3452–3462, <http://dx.doi.org/10.1016/j.neuroimage.2009.11.012>.
- Rothmayr, C., Sodian, B., Hajak, G., Döhnel, K., Meinhart, J., Sommer, M., 2011. Common and distinct neural networks for false-belief reasoning and inhibitory control. *Neuroimage* 56, 1705–1713, <http://dx.doi.org/10.1016/j.neuroimage.2010.12.052>.
- Sallet, J., Mars, R.B., Noonan, M.P., Neubert, F.X., Jbabdi, S., O'Reilly, J.X., Filippini, N., Thomas, A.G., Rushworth, M.F., 2013. The organization of dorsal frontal cortex in humans and macaques. *Journal of Neuroscience* 33, 12255–12274, <http://dx.doi.org/10.1523/JNEUROSCI.5108-12.2013>.
- Sarfati, Y., Hardy-Baylé, M.C., Besche, C., Widlocher, D., 1997. Attribution of intentions to others in people with schizophrenia: a non-verbal exploration with comic strips. *Schizophrenia Research* 25, 199–209.
- Saxe, R., 2006. Uniquely human social cognition. *Current Opinion in Neurobiology* 16, 235–239.
- Saxe, R., Carey, S., Kanwisher, N., 2004a. Understanding other minds: linking developmental psychology and functional neuroimaging. *Annual Review of Psychology* 55, 87–124.
- Saxe, R., Kanwisher, N., 2003. People thinking about thinking people. The role of the temporo-parietal junction in "theory of mind". *Neuroimage* 19, 1835–1842.
- Saxe, R., Powell, L.J., 2006. It's the thought that counts: specific brain regions for one component of theory of mind. *Psychological Science* 17, 692–699.
- Saxe, R., Xiao, D.K., Kovacs, G., Perrett, D.I., Kanwisher, N., 2004b. A region of right posterior superior temporal sulcus responds to observed intentional actions. *Neuropsychologia* 42, 1435–1446.
- Saygin, A.P., 2007. Superior temporal and premotor brain areas necessary for biological motion perception. *Brain* 130, 2452–2461.
- Scheibel, R.S., Newsome, M.R., Wilde, E.A., McClelland, M.M., Hanten, G., Krawczyk, D.C., Cook, L.G., Chu, Z.D., Vásquez, A.C., Yallampalli, R., Lin, X., Hunter, J.V., Levin, H.S., 2011. Brain activation during a social attribution task in adolescents with moderate to severe traumatic brain injury. *Social Neuroscience* 6, 582–598, <http://dx.doi.org/10.1080/17470919.2011.588844>.
- Scholz, J., Triantafyllou, C., Whitfield-Gabrieli, S., Brown, E.N., Saxe, R., 2009. Distinct regions of right temporo-parietal junction are selective for the theory of mind and exogenous attention. *PLoS One* 4, e4869, <http://dx.doi.org/10.1371/journal.pone.0004869>.
- Schroeter, M.L., Ettrich, B., Menz, M., Zysset, S., 2010. Traumatic brain injury affects the frontomedian cortex—an event-related fMRI study on evaluative judgments. *Neuropsychologia* 48, 185–193, <http://dx.doi.org/10.1016/j.neuropsychologia.2009.09.004>.
- Schulte-Rüther, M., Greimel, E., Markowitsch, H.J., Kamp-Becker, I., Remschmidt, H., Fink, G.R., Piefke, M., 2011. Dysfunctions in brain networks supporting empathy: an fMRI study in adults with autism spectrum disorders. *Social Neuroscience* 6, 1–21, <http://dx.doi.org/10.1080/17470911003708032>.
- Schultz, R.T., Grelotti, D.J., Klein, A., Kleinman, J., Van der Gaag, C., Marois, R., Skudlarski, P., 2003. The role of the fusiform face area in social cognition: implications for the pathobiology of autism. *Philosophical Transactions of the Royal Society B: Biological Sciences* 358, 415–427.
- Schuwerk, T., Döhnel, K., Sodian, B., Keck, I. R., Rupprecht, R. and Sommer, M. Functional activity and effective connectivity of the posterior medial prefrontal cortex during processing of incongruent mental states. *Human Brain Mapping*, in press, doi: 10.1002/hbm.22377.
- Shamay-Tsoory, S.G., Aharon-Peretz, J., Perry, D., 2009. Two systems for empathy: a double dissociation between emotional and cognitive empathy in inferior frontal gyrus versus ventromedial prefrontal lesions. *Brain* 132, 617–627, <http://dx.doi.org/10.1093/brain/awn279>.
- Sommer, M., Meinhardt, J., Eichenmüller, K., Sodian, B., Döhnel, K., Hajak, G., 2010. Modulation of the cortical false belief network during development. *Brain Research* 1354, 123–131, <http://dx.doi.org/10.1016/j.brainres.2010.07.057>.
- Spreng, R.N., Mar, R.A., Kim, A.S., 2009. The common neural basis of autobiographical memory, prospection, navigation, theory of mind, and the default mode: a quantitative meta-analysis. *Journal of Cognitive Neuroscience* 21, 489–510, <http://dx.doi.org/10.1162/jocn.2008.21029>.
- Stone, V.E., Baron-Cohen, S., Knight, R.T., 1998. Frontal lobe contributions to theory of mind. *Journal of Cognitive Neuroscience* 10, 640–656.
- Straube, B., Green, A., Bromberger, B., Kircher, T., 2011. The differentiation of iconic and metaphoric gestures: common and unique integration processes. *Human Brain Mapping* 32, 520–533, <http://dx.doi.org/10.1002/hbm.21041>.
- Sugiura, M., Sassa, Y., Watanabe, J., Akitsuki, Y., Maeda, Y., Matsue, Y., Fukuda, H., Kawashima, R., 2006. Cortical mechanisms of person representation: recognition of famous and personally familiar names. *Neuroimage* 31, 853–860.
- Suchan, B., Yáñez, L., Wunderlich, G., Canavan, A.G., Herzog, H., Tellmann, L., Hömberg, V., Seitz, R.J., 2002. Hemispheric dissociation of visual-pattern processing and visual rotation. *Behavioral Brain Research* 136, 533–544.
- Tager-Flusberg, H., Boshart, J., Baron-Cohen, S., 1998. Reading the windows of the soul: evidence of domain specificity sparing in Williams syndrome. *Journal of Cognitive Neuroscience* 10, 631–639.
- Tamir, D.I., Mitchell, J.P., 2010. Neural correlates of anchoring-and-adjustment during mentalizing. *Proceedings of the National Academy of Sciences of the United States of America* 107, 10827–10832, <http://dx.doi.org/10.1073/pnas.1003242107>.
- Tomlin, D., Kayali, M.A., King-Casas, B., Anen, C., Camerer, C.F., Quartz, S.R., Montague, P.R., 2006. Agent-specific responses in the cingulate cortex during economic exchanges. *Science* 312, 1047–1050.
- Tsukiura, T., Mano, Y., Sekiguchi, A., Yomogida, Y., Hoshi, K., Kambara, T., Takeuchi, H., Sugiura, M., Kawashima, R., 2010. Dissociable roles of the anterior temporal regions in successful encoding of memory for person identity information. *Journal of Cognitive Neuroscience* 22, 2226–2237.
- Vanderwal, T., Hunyadi, E., Gruppe, D.W., Connors, C.M., Schultz, R.T., 2008. Self, mother and abstract other: an fMRI study of reflective social processing. *Neuroimage* 41, 1437–1446, <http://dx.doi.org/10.1016/j.neuroimage.2008.03.058>.
- Van Overwalle, F., 2009. Social cognition and the brain: a meta-analysis. *Human Brain Mapping* 30, 829–858, <http://dx.doi.org/10.1002/hbm.20547>.
- Van Overwalle, F., Baetens, K., 2009. Understanding others' actions and goals by mirror and mentalizing systems: a meta-analysis. *Neuroimage* 48, 564–584, <http://dx.doi.org/10.1016/j.neuroimage.2009.06.009>.
- Vogeley, K., May, M., Ritzl, A., Falkai, P., Zilles, K., Fink, G.R., 2004. Neural correlates of first-person perspective as one constituent of human self-consciousness. *Journal of Cognitive Neuroscience* 16, 817–827.
- Wager, T.D., Lindquist, M., Kaplan, L., 2007. Meta-analysis of functional neuroimaging data: current and future directions. *Social Cognitive and Affective Neuroscience* 2, 150–158, <http://dx.doi.org/10.1093/scan/nsm015>.
- Yoshida, W., Seymour, B., Friston, K.J., Dolan, R.J., 2010. Neural mechanisms of belief inference during cooperative games. *Journal of Neuroscience* 30, 10744–10751, <http://dx.doi.org/10.1523/JNEUROSCI.5895-09.2010>.
- Zahn, R., Moll, J., Krueger, F., Huey, E., Garrido, G., Grafman, J., 2007. Social concepts are represented in the superior anterior temporal cortex. *Proceedings of the National Academy of Sciences of the United States of America* 104, 6430–6435.
- Zaitchik, D., 1990. When representations conflict with reality: the preschooler's problem with false beliefs and "false" photographs. *Cognition* 35, 41–68.

## References in the meta-analysis

- Abraham, A., Rakoczy, H., Werning, M., von Cramon, D.Y., Schubotz, R.I., 2010. Matching mind to world and vice versa: functional dissociations between belief and desire mental state processing. *Social Neuroscience* 5, 1–18, <http://dx.doi.org/10.1080/17470910903166853>.
- Aichhorn, M., Perner, J., Weiss, B., Kronbichler, M., Staffen, W., Ladurner, G., 2009. Temporo-parietal junction activity in theory-of-mind tasks: false-self beliefs, or attention. *Journal of Cognitive Neuroscience*, 1179–1192, <http://dx.doi.org/10.1162/jocn.2009.21082>.
- Adams Jr., R.B., Rule, N.O., Franklin Jr., R.G., Wang, E., Stevenson, M.T., Yoshikawa, S., Nomura, M., Sato, W., Kveraga, K., Ambady, N., 2010. Cross-cultural reading the mind in the eyes: an fMRI investigation. *Journal of Cognitive Neuroscience* 22, 97–108, <http://dx.doi.org/10.1162/jocn.2009.21187>.
- Assaf, M., Kahn, I., Pearlson, G.D., Johnson, M.R., Yeshurun, Y., Calhoun, V.D., Helder, T., 2009. Brain activity dissociates mentalization from motivation during an interpersonal competitive game. *Brain Imaging Behavior* 3, 24–37.
- Bahnemann, M., Dzibek, I., Prehn, K., Wolf, I., Heeckeren, H.R., 2010. Sociotopy in the temporoparietal cortex: common versus distinct processes. *Social Cognitive and Affective Neuroscience* 5, 48–58, <http://dx.doi.org/10.1093/scan/nsp045>.
- Baron-Cohen, S., Ring, H.A., Wheelwright, S., Bullmore, E.T., Brammer, M.J., Simmons, A., Williams, S.C., 1999. Social intelligence in the normal and autistic brain: an fMRI study. *European Journal of Neuroscience* 11, 1891–1898.
- Blakemore, S.J., Boyer, P., Pachot-Clouard, M., Meltzoff, A., Segebarth, C., Decety, J., 2003. The detection of contingency and animacy from simple animations in the human brain. *Cerebral Cortex* 13, 837–844.
- Bruene, M., Lissek, S., Fuchs, N., Witthaus, H., Peters, S., Nicolas, V., Juckel, G., Tegenthoff, M., 2008. An fMRI study of theory of mind in schizophrenic patients with "passivity" symptoms. *Neuropsychologia* 46, 1992–2001, <http://dx.doi.org/10.1016/j.neuropsychologia.2008.01.023>.
- Brunet, E., Sarfati, Y., Hardy-Baylé, M.C., Decety, J., 2000. A PET investigation of the attribution of intentions with a nonverbal task. *Neuroimage* 11, 157–166.
- Castelli, F., Happé, F., Frith, U., Frith, C.D., 2000. Movement and mind: a functional imaging study of perception and interpretation of complex intentional movement patterns. *Neuroimage* 12, 314–325.
- Castelli, I., Baglio, F., Blasi, V., Alberoni, M., Falini, A., Liverta-Sempio, O., Nemni, R., Marchetti, A., 2010. Effects of aging on mindreading ability through the eyes: an fMRI study. *Neuropsychologia* 48, 2586–2594, <http://dx.doi.org/10.1016/j.neuropsychologia.2010.05.005>.
- Chaminade, T., Rosset, D., Da Fonseca, D., Nazarian, B., Lutcher, E., Cheng, G., Deruelle, C., 2012. How do we think machines think? An fMRI study of alleged competition with an artificial intelligence. *Frontiers in Human Neuroscience* 6, 103, <http://dx.doi.org/10.3389/fnhum.2012.00103>.
- Craik, F.I.M., Moroz, T.M., Moscovitch, M., Stuss, D.T., Winocur, G., Tulving, E., Kapur, S., 1999. In search of the self: a positron emission tomography study. *Psychological Science* 10, 26–34, <http://dx.doi.org/10.1111/1467-9280.00102>.
- Das, P., Lagopoulos, J., Coulston, C.M., Henderson, A.F., Malhi, G.S., 2012. Mentalizing impairment in schizophrenia: a functional MRI study. *Schizophrenia Research* 134, 158–164, <http://dx.doi.org/10.1016/j.schres.2011.08.019>.
- Decety, J., Jackson, P.L., Sommerville, J.A., Chaminade, T., Meltzoff, A.N., 2004. The neural bases of cooperation and competition: an fMRI investigation. *Neuroimage* 23, 744–751.
- De Achával, D., Villareal, M.F., Costanzo, E.Y., Douer, J., Castro, M.N., Mora, M.C., Nemeroff, C.B., Chu, E., Bär, K.J., Guinjoan, S.M., 2012. Decreased activity in right-hemispheric structures involved in social cognition in siblings discordant for schizophrenia. *Schizophrenia Research* 134, 171–179, <http://dx.doi.org/10.1016/j.schres.2011.11.010>.
- Dodell-Feder, D., Koster-Hale, J., Bedny, M., Saxe, R., 2011. fMRI item analysis in a theory of mind task. *Neuroimage* 55, 705–712, <http://dx.doi.org/10.1016/j.neuroimage.2010.12.040>.
- Fletcher, P.C., Happé, F., Frith, U., Baker, S.C., Dolan, R.J., Frackowiak, R.S., Frith, C.D., 1995. Other minds in the brain: a functional imaging study of "theory of mind" in story comprehension. *Cognition* 57, 109–128.
- Focquaert, F., Steven-Wheeler, M.S., Vanneste, S., Doron, K.W., Platek, S.M., 2010. Mindreading in individuals with an empathizing versus systemizing cognitive style: an fMRI study. *Brain Research Bulletin* 83, 214–222, <http://dx.doi.org/10.1016/j.brainresbulletin.2010.08.008>.
- Fukui, H., Murai, T., Shinozaki, J., Aso, T., Fukuyama, H., Hayashi, T., Hanakawa, T., 2006. The neural basis of social tactics: an fMRI study. *Neuroimage* 32, 913–920.
- Gallagher, H.L., Frith, C.D., 2004. Dissociable neural pathways for the perception and recognition of expressive and instrumental gestures. *Neuropsychologia* 42, 1725–1736.
- Gallagher, H.L., Happé, 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, 11–21.
- Gallagher, H.L., Jack, A.I., Roepstorff, A., Frith, C.D., 2002. Imaging the intentional stance in a competitive game. *Neuroimage* 16, 814–821.
- Gobbini, M.I., Koralek, A.C., Bryan, R.E., Montgomery, K.J., Haxby, J.V., 2007. Two takes on the social brain: a comparison of theory of mind tasks. *Journal of Cognitive Neuroscience* 19, 1803–1814.
- Happé, F., Ehlers, S., Fletcher, P., Frith, U., Johansson, M., Gillberg, C., Dolan, R., Frackowiak, R., Frith, C.D., 1996. Theory of mind in the brain. Evidence from a PET scan study of Asperger syndrome. *Neuroreport* 8, 197–201.
- Heatherton, T.F., Wyland, C.L., Macrae, C.N., Demos, K.E., Denny, B.T., Kelley, W.M., 2006. Medial prefrontal activity differentiates self from close others. *Social Cognitive and Affective Neuroscience* 1, 18–25, <http://dx.doi.org/10.1093/scan/nsl001>.
- Hynes, C.A., Baird, A.A., Grafton, S.T., 2006. Differential role of the orbital frontal lobe in emotional versus cognitive perspective-taking. *Neuropsychologia* 44, 374–383.
- Jimura, K., Konishi, S., Asari, T., Miyashita, Y., 2010. Temporal pole activity during understanding other persons' mental states correlates with neuroticism trait. *Brain Research* 1328, 104–112, <http://dx.doi.org/10.1016/j.brainres.2010.03.016>.
- Kana, R.K., Keller, T.A., Cherkassky, V.L., Minshew, N.J., Just, M.A., 2009. Atypical frontal-posterior synchronization of theory of mind regions in autism during mental state attribution. *Social Neuroscience* 4, 135–152, <http://dx.doi.org/10.1080/17470910802198510>.
- Kircher, T., Bluemel, I., Marjoram, D., Lataster, T., Krabbendam, L., Weber, J., van Os, J., Krach, S., 2009. Online mentalising investigated with functional MRI. *Neuroscience Letters* 454, 176–181, <http://dx.doi.org/10.1016/j.neulet.2009.03.026>.
- Kliemann, D., Young, L., Scholz, J., Saxe, R., 2008. The influence of prior record on moral judgment. *Neuropsychologia* 46, 2949–2957, <http://dx.doi.org/10.1016/j.neuropsychologia.2008.06.010>.
- Kobayashi, C., Glover, G.H., Temple, E., 2006. Cultural and linguistic influence on neural bases of 'Theory of Mind': an fMRI study with Japanese bilinguals. *Brain and Language* 98, 210–220.
- Kobayashi, C., Glover, G.H., Temple, E., 2007. Children's and adults' neural bases of verbal and nonverbal 'theory of mind'. *Neuropsychologia* 45, 1522–1532.
- Koelkebeck, K., Hirao, K., Kawada, R., Miyata, J., Saze, T., Ubukata, S., Itakura, S., Kanakogi, Y., Ohrrmann, P., Bauer, J., Pedersen, A., Sawamoto, N., Fukuyama, H., Takahashi, H., Murai, T., 2011. Transcultural differences in brain activation patterns during theory of mind (ToM) task performance in Japanese and Caucasian participants. *Social Neuroscience* 6, 615–626, <http://dx.doi.org/10.1080/17470919.2011.620763>.
- Krach, S., Bluemel, I., Marjoram, D., Lataster, T., Krabbendam, L., Weber, J., van Os, J., Kircher, T., 2009. Are women better mindreaders? Sex differences in neural correlates of mentalizing detected with functional MRI. *BMC Neuroscience* 4 (10), 9, <http://dx.doi.org/10.1186/1471-2202-10-9>.
- Krueger, F., McCabe, K., Moll, J., Kriegeskorte, N., Zahn, R., Strenziok, M., Heinecke, A., Grafman, J., 2007. Neural correlates of trust. *Proceedings of the National Academy of Sciences of the United States of America* 104, 20084–20089.
- Lee, J., Quintana, J., Nori, P., Green, M.F., 2011. Theory of mind in schizophrenia: exploring neural mechanisms of belief attribution. *Social Neuroscience* 6, 569–581, <http://dx.doi.org/10.1080/17470919.2011.620774>.
- Lou, H.C., Luber, B., Crupain, M., Keenan, J.P., Nowak, M., Kjaer, T.W., Sackeim, H.A., Lisanby, S.H., 2004. Parietal cortex and representation of the mental self. *Proceedings of the National Academy of Sciences of the United States of America* 101, 6827–6832.
- Ma, N., Vanderkerckhove, M., Van Overwalle, F., Seurinck, R., Fias, W., 2011. Spontaneous and intentional trait inferences recruit a common mentalizing network to a different degree: spontaneous inferences activate only its core areas. *Social Neuroscience* 6, 123–138, <http://dx.doi.org/10.1080/17470919.2010.485884>.
- Malhi, G.S., Lagopoulos, J., Das, P., Moss, K., Berk, M., Coulston, C.M., 2008. A functional MRI study of Theory of Mind in euthymic bipolar disorder patients. *Bipolar Disorders* 10, 943–956, <http://dx.doi.org/10.1111/j.1399-5618.2008.00643.x>.
- Martin, A., Weisberg, J., 2003. Neural foundations for understanding social and mechanical concepts. *Cognitive Neuropsychology* 20, 575–587.
- McAdams, C.J., Krawczyk, D.C., 2011. Impaired neural processing of social attribution in anorexia nervosa. *Psychiatry Research* 194, 54–63, <http://dx.doi.org/10.1016/j.psychres.2011.06.016>.
- Mitchell, J.P., 2008. Activity in right temporo-parietal junction is not selective for theory-of-mind. *Cerebral Cortex* 18, 262–271.
- Mitchell, J.P., Banaji, M.R., Macrae, C.N., 2005a. General and specific contributions of the medial prefrontal cortex to knowledge about mental states. *Neuroimage* 28, 757–762.
- Mitchell, J.P., Banaji, M.R., Macrae, C.N., 2005b. The link between social cognition and self-referential thought in the medial prefrontal cortex. *Journal of Cognitive Neuroscience* 17, 1306–1315.
- Mitchell, J.P., Cloutier, J., Banaji, M.R., Macrae, C.N., 2006. Medial prefrontal dissociations during processing of trait diagnostic and nondiagnostic person information. *Social Cognitive and Affective Neuroscience* 1, 49–55, <http://dx.doi.org/10.1093/scan/nsl007>.
- Mitchell, J.P., Heatherton, T.F., Macrae, C.N., 2002. Distinct neural systems subserve person and object knowledge. *Proceedings of the National Academy of Sciences of the United States of America* 99, 15238–15243.
- Mitchell, J.P., Macrae, C.N., Banaji, M.R., 2004. Encoding-specific effects of social cognition on the neural correlates of subsequent memory. *Journal of Neuroscience* 24, 4912–4917.
- Mitchell, J.P., Macrae, C.N., Banaji, M.R., 2005c. Forming impressions of people versus inanimate objects: social-cognitive processing in the medial prefrontal cortex. *Neuroimage* 26, 251–257.
- Modinos, G., Renken, R., Ormel, J., Aleman, A., 2011. Self-reflection and the psychosis-prone brain: an fMRI study. *Neuropsychology* 25, 295–305, <http://dx.doi.org/10.1037/a0021747>.
- Moor, B.G., Mack, Z.A., Gueroglu, B., Rombouts, S.A., Molen, M.W., Crone, E.A., 2012. Neurodevelopmental changes of reading the mind in the eyes. *Social Cognitive and Affective Neuroscience* 7, 44–52, <http://dx.doi.org/10.1093/scan/nsr020>.
- Moriguchi, Y., Ohnishi, T., Lane, R.D., Maeda, M., Mori, T., Nemoto, K., Matsuda, H., Komaki, G., 2006. Impaired self-awareness and theory of mind: an fMRI study of mentalizing in alexithymia. *Neuroimage* 32, 1472–1482.

- Murphy, E.R., Brent, B.K., Benton, M., Pruitt, P., Diwadkar, V., Rajarethnam, R.P., Keshavan, M.S., 2010. Differential processing of metacognitive evaluation and the neural circuitry of the self and others in schizophrenia: a pilot study. *Schizophrenia Research* 116, 252–258, <http://dx.doi.org/10.1016/j.schres.2009.11.009>.
- Nieminen-von Wendt, T., Metsähonkala, L., Kulomäki, T., Aalto, S., Autti, T., Vanhala, R., von Wendt, L., 2003. Changes in cerebral blood flow in Asperger syndrome during theory of mind tasks presented by the auditory route. *European Child & Adolescent Psychiatry* 12, 178–189.
- Ochsner, K.N., Beer, J.S., Robertson, E.R., Cooper, J.C., Gabrieli, J.D., Kihlstrom, J.F., D'Esposito, M., 2005. The neural correlates of direct and reflected self-knowledge. *NeuroImage* 28, 797–814.
- Perner, J., Aichhorn, M., Kronbichler, M., Staffen, W., Ladurner, G., 2006. Thinking of mental and other representations: the roles of left and right temporo-parietal junction. *Social Neuroscience* 1, 245–258, <http://dx.doi.org/10.1080/17470910600989896>.
- Platek, S.M., Keenan, J.P., Gallup Jr., G.G., Mohamed, F.B., 2004. Where am I? The neurological correlates of self and other. *Cognitive Brain Research* 19, 114–122.
- Rilling, J.K., Dagenais, J.E., Goldsmith, D.R., Glenn, A.L., Pagoni, G., 2008. Social cognitive neural networks during in-group and out-group interactions. *NeuroImage* 41, 1447–1461, <http://dx.doi.org/10.1016/j.neuroimage.2008.03.044>.
- Rilling, J.K., Sanfey, A.G., Aronson, J.A., Nystrom, L.E., Cohen, J.D., 2004. The neural correlates of theory of mind within interpersonal interactions. *NeuroImage* 22, 1694–1703.
- Ross, L.A., Olson, I.R., 2010. Social cognition and the anterior temporal lobes. *NeuroImage* 49, 3452–3462, <http://dx.doi.org/10.1016/j.neuroimage.2009.11.012>.
- Russell, T.A., Rubia, K., Bullmore, E.T., Soni, W., Suckling, J., Brammer, M.J., Simmons, A., Williams, S.C., Sharma, T., 2000. Exploring the social brain in schizophrenia: left prefrontal underactivation during mental state attribution. *American Journal of Psychiatry* 157, 2040–2042.
- Samson, D., Apperly, I.A., Chiavarino, C., Humphreys, G.W., 2004. Left temporo-parietal junction is necessary for representing someone else's belief. *Nature Neuroscience* 7, 499–500.
- Santos, N.S., Kuzmanovic, B., David, N., Rotarska-Jagiela, A., Eickhoff, S.B., Shah, J.N., Fink, G.R., Bente, G., Vogeley, K., 2010. Animated brain: a functional neuroimaging study on animacy experience. *NeuroImage* 53, 291–302, <http://dx.doi.org/10.1016/j.neuroimage.2010.05.080>.
- Saxe, R., Kanwisher, N., 2003. People thinking about thinking people. The role of the temporo-parietal junction in "theory of mind". *NeuroImage* 19, 1835–1842.
- Saxe, R., Powell, L.J., 2006. It's the thought that counts: specific brain regions for one component of theory of mind. *Psychological Science* 17, 692–699.
- Saxe, R., Schulz, L.E., Jiang, Y.V., 2006. Reading minds versus following rules: dissociating theory of mind and executive control in the brain. *Social Neuroscience* 1, 284–298, <http://dx.doi.org/10.1080/1747091061000446>.
- Saxe, R., Wexler, A., 2005. Making sense of another mind: the role of the right temporo-parietal junction. *Neuropsychologia* 43, 1391–1399.
- Schmitz, T.W., Kawahara-Baccus, T.N., Johnson, S.C., 2004. Metacognitive evaluation, self-relevance, and the right prefrontal cortex. *NeuroImage* 22, 941–947.
- Schnell, K., Bluschke, S., Konradt, B., Walter, H., 2011. Functional relations of empathy and mentalizing: an fMRI study on the neural basis of cognitive empathy. *NeuroImage* 54, 1743–1754, <http://dx.doi.org/10.1016/j.neuroimage.2010.08.024>.
- Schultz, J., Imamizu, H., Kawato, M., Frith, C.D., 2004. Activation of the human superior temporal gyrus during observation of goal attribution by intentional objects. *Journal of Cognitive Neuroscience* 16, 1695–1705.
- Sebastian, C.L., Fontaine, N.M., Bird, G., Blakemore, S.J., Brito, S.A., McCrory, E.J., Viding, E., 2012. Neural processing associated with cognitive and affective Theory of Mind in adolescents and adults. *Social Cognitive and Affective Neuroscience* 7, 52–63, <http://dx.doi.org/10.1093/scan/nsr023>.
- Sripada, C.S., Angstadt, M., Banks, S., Nathan, P.J., Liberzon, I., Phan, K.L., 2009. Functional neuroimaging of mentalizing during the trust game in social anxiety disorder. *NeuroReport* 20, 984–989, <http://dx.doi.org/10.1097/WNR.0b013e32832d0a67>.
- Suzuki, S., Niki, K., Fujisaki, S., Akiyama, E., 2011. Neural basis of conditional cooperation. *Social Cognitive and Affective Neuroscience* 6, 338–347, <http://dx.doi.org/10.1093/scan/nsq042>.
- Tavares, P., Lawrence, A.D., Barnard, P.J., 2008. Paying attention to social meaning: an fMRI study. *Cerebral Cortex* 18, 1876–1885.
- Todorov, A., Gobbini, M.I., Evans, K.K., Haxby, J.V., 2007. Spontaneous retrieval of affective person knowledge in face perception. *Neuropsychologia* 45, 163–173.
- Villareal, M.F., Fridman, E.A., Leiguarda, R.C., 2012. The effect of the visual context in the recognition of symbolic gestures. *PLoS One* 7, e29644, <http://dx.doi.org/10.1371/journal.pone.0029644>.
- Voellm, B.A., Taylor, A.N., Richardson, P., Corcoran, R., Stirling, J., McKie, S., Deakin, J.F., Elliott, R., 2006. Neuronal correlates of theory of mind and empathy: a functional magnetic resonance imaging study in a nonverbal task. *NeuroImage* 29, 90–98.
- Walter, H.W., Adenzato, M., Ciaramidaro, A., Enrici, I., Lorenzo, P., Bara, B.G., 2004. Understanding intentions in social interaction: the role of the anterior paracingulate cortex. *Journal of Cognitive Neuroscience* 16, 1854–1863.
- Walter, H.W., Ciaramidaro, A., Adenzato, M., Vasic, N., Ardito, R.B., Erk, S., Bara, B.G., 2009. Dysfunction of the social brain in schizophrenia is modulated by intention type: an fMRI study. *Social Cognitive and Affective Neuroscience* 4, 166–176, <http://dx.doi.org/10.1093/scan/nsn047>.
- Wicker, B., Perrett, D.I., Baron-Cohen, S., Decety, J., 2003. Being the target of another's emotion: a PET study. *Neuropsychologia* 41, 139–146.
- Young, L., Camprodon, J.A., Hauser, M., Pascual-Leone, A., Saxe, R., 2010a. Disruption of the right temporo-parietal junction with transcranial magnetic stimulation reduces the role of beliefs in moral judgments. *Proceedings of the National Academy of Sciences of the United States of America* 107, 6753–6758.
- Young, L., Cushman, F., Hauser, M., Saxe, R., 2007. The neural basis of the interaction between theory of mind and moral judgment. *Proceedings of the National Academy of Sciences of the United States of America* 104, 8235–8240.
- Young, L., Dodell-Feder, D., Saxe, R., 2010b. What gets the attention of the temporo-parietal junction? An fMRI investigation of attention and theory of mind. *Neuropsychologia* 48, 2658–2664, <http://dx.doi.org/10.1016/j.neuropsychologia.2010.05.012>.
- Young, L., Saxe, R., 2009. An fMRI investigation of spontaneous mental state inference for moral judgment. *Journal of Cognitive Neuroscience* 21, 1396–1405, <http://dx.doi.org/10.1162/jocn.2009.21137>.
- Young, L., Scholz, J., Saxe, R., 2011. Neural evidence for "intuitive prosecution": the use of mental state information for negative moral verdicts. *Social Neuroscience* 6, 302–315, <http://dx.doi.org/10.1080/17470919.2010.529712>.
- Zhu, Y., Zhang, L., Fan, J., Han, S., 2007. Neural basis of cultural influence on self-representation. *NeuroImage* 34, 1310–1316.