NeuroImage 61 (2012) 921–930

Contents lists available at SciVerse ScienceDirect

NeuroImage

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

Multiple roles for executive control in belief–desire reasoning: Distinct neural networks are recruited for self perspective inhibition and complexity of reasoning

Charlotte E. Hartwright [*], Ian A. Apperly, Peter C. Hansen

University of Birmingham, UK

ARTICLE INFO

Article history:
Accepted 5 March 2012
Available online 14 March 2012

Keywords:
Theory of mind
Perspective taking
Executive control
Inhibition
Inferior frontal gyrus
Ventrolateral prefrontal cortex

ABSTRACT

Belief–desire reasoning is a core component of 'Theory of Mind' (ToM), which can be used to explain and predict the behaviour of agents. Neuroimaging studies reliably identify a network of brain regions comprising a 'standard' network for ToM, including temporoparietal junction and medial prefrontal cortex. Whilst considerable experimental evidence suggests that executive control (EC) may support a functioning ToM, coordination of neural systems for ToM and EC is poorly understood. We report here use of a novel task in which psychologically relevant ToM parameters (true versus false belief; approach versus avoidance desire) were manipulated orthogonally. The valence of these parameters not only modulated brain activity in the 'standard' ToM network but also in EC regions. Varying the valence of both beliefs and desires recruits anterior cingulate cortex, suggesting a shared inhibitory component associated with negatively valenced mental state concepts. Varying the valence of beliefs additionally draws on ventrolateral prefrontal cortex, reflecting the need to inhibit self perspective. These data provide the first evidence that separate functional and neural systems for EC may be recruited in the service of different aspects of ToM.

© 2012 Elsevier Inc. All rights reserved.

Introduction

The capacity to reason about the mental causes of action, termed 'mentalising' or exercising a 'Theory of Mind' (ToM), has received considerable interest from social neuroscientists over the last decade. Much attention has been given to identifying which, if any, brain regions should be considered as specialised for ToM. This work has made considerable progress in identifying possible contenders, and converges on the importance of a network of brain regions including temporoparietal junction (TPJ) and medial prefrontal cortex (mPFC) (Carrington and Bailey, 2009; Lieberman, 2007; Mar, 2011; van Overwalle, 2009). TPJ has been identified in the majority of neuroimaging studies of ToM, and appears selectively responsive when representing mental states such as beliefs, desires and intentions, over and above representation of physical states, personality traits or dispositions of the person, and above non-mental representations, such as photographs (Aichorn et al., 2009; Saxe and Kanwisher, 2003; Saxe and Powell, 2006; Saxe and Wexler, 2005). mPFC is also identified in most neuroimaging studies of ToM, though its activity may be less specific to mental state representation (Amodio and Frith, 2006), and may be most strongly recruited when reflecting on more enduring mental states, such as personality traits and social or moral beliefs (van Overwalle, 2009), or when making inferences under conditions of high uncertainty (Jenkins and Mitchell, 2010). The strong convergence of neuroimaging data has lead to a general consensus that TPJ and mPFC constitute the 'core' network for ToM, and that the functions they support are the most psychologically important for understanding ToM.

ToM has been studied most extensively using false belief tasks. A classic paradigm, the object transfer task, requires participants to make a prediction about the behaviour of a character, based upon the character's belief and desire at that point in time. A typical experimental sequence outlines a protagonist putting an object into location A. They then leave the scene. Whilst the protagonist is away, the object is transferred to location B. The character then returns, wishing to find the object but holding a false belief about its location (Wimmer and Perner, 1983). In order to successfully identify where the protagonist will look for the object, it is necessary for participants to infer the character's false belief about the object's location and predict the character's action on the basis of their false belief, whilst resisting interference from their own privileged knowledge of the object's true location, and what the right course of action would be. This task analysis leads to the expectation that successful ToM will not only require processes that might be specific to inferring and representing the mental states of others, but also processes for executive control (EC) to ensure that the correct information is selected for inferring mental states and predicting actions. It follows, then, that a complete account of the neural basis of ToM must also include brain regions associated with these sorts of control processes. To date, however, the brain bases of EC in ToM have been little explored.

[*] Corresponding author at: Centre for Behavioural Brain Sciences, 2.30 Hills Building, School of Psychology, University of Birmingham, B15 2TT, UK. Fax: +44 121 414 4897.
E-mail address: cee849@bham.ac.uk (C.E. Hartwright).

1053-8119/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2012.03.012

Numerous researchers have noted that executive ability appears to contribute significantly to proficiency with ToM (e.g., Carlson and Moses, 2001; Carlson et al., 1998, 2002; Friedman and Leslie, 2004, 2005; German and Hehman, 2006; Leslie and Polizzi, 1998; Leslie et al., 2005; Perner and Lang, 1999; Wellman et al., 2001). For example, in the classic false belief paradigm mentioned above, children under the age of four seem unable to overcome their own knowledge of where the object really is. As a result, they consistently state that the protagonist will look for the object in the object's true location. Younger children, however, may sometimes pass the false belief task if the true location of the object is made less salient (Carlson et al., 1998; Wellman et al., 2001). These kind of 'egocentric errors', sometimes referred to as the 'curse of knowledge' (Birch and Bloom, 2004, 2007) or a 'reality bias' (Mitchell and Lacohee, 1991), have also been observed in older children and healthy adults (Bernstein et al., 2004; Birch and Bloom, 2007), and appear to reflect the need to exert EC to solve such tasks (see e.g., Apperly et al., 2005, 2009, for relevant discussions).

Two broad theoretical frameworks have been proposed concerning the role of EC in ToM. The first suggests that EC is necessary when a perspective difference between self and other exists, as is the case of false belief or conflicting desires. For example, knowledge of the true location may interfere with the ability to select the believed, or false, location. As a result, the self perspective must be inhibited in order to assume the perspective of the other (Ruby and Decety, 2003; Samson et al., 2005). This theory arises from behavioural observations of young children's propensity towards responding with their own knowledge, and data suggesting that performance in ToM tasks can be manipulated by varying the salience of self perspective (Carlson et al., 1998; Wellman et al., 2001). A growing literature suggests that the ventrolateral prefrontal cortex (vlPFC) may support this process of 'self perspective inhibition'. For example, Vogeley et al. (2001) identified that the right inferior frontal cortex, particularly right inferior frontal gyrus (rIFG), was modulated by varying the importance of self in a fictional scenario. This finding was later supported by a case study which demonstrated that damage to right vlPFC, including rIFG, resulted in interference from self perspective when attributing beliefs to others. In this particular case, the patient was able to solve ToM tasks where his own perspective was less salient, but failed ToM tasks where a clear incongruence between self and other knowledge state existed (Samson et al., 2005). Using false belief tasks from Samson et al.'s study and a stop-signal test of EC, a further study showed that the same ventral region of IFG was recruited bilaterally in healthy adults for both general response inhibition, and when contrasting false belief tasks that made high versus low demands on the inhibition of self-perspective (van der Meer et al., 2011). Finally, a recent study of visual perspective-taking showed an ERP component over right fronto-lateral cortex that was sensitive to differences between self and other perspectives (McCleery et al., 2011). These studies provide converging evidence that a functioning ToM is supported by regions outside of the 'standard' ToM network, and that the inferior frontal cortex – particularly vlPFC – may be an important, but overlooked, region involved in inhibition of self perspective. Notably, however, none of these studies examines the role of EC in reasoning about conative mental states, such as desires.

The second theory of the role of EC in ToM, proposed by Leslie and colleagues (Friedman and Leslie, 2004, 2005; Leslie and Polizzi, 1998; Leslie et al., 2005), extends beyond belief attribution to include the varying demands of desire reasoning. It is implicit in the standard false belief task that the character wishes to locate the object. However, if the agent holds a desire to avoid the object, both children and adults suffer further difficulty in false belief tasks (Apperly et al., 2011; Cassidy, 1998; German and Hehman, 2006). Moreover, like false belief reasoning, proficiency with avoidance desire coincides with the development of executive abilities. Leslie and colleagues explain this in terms of a shared inhibitory component for negatively valenced[1] mental states, such as false belief and avoidance desire. They suggest that, for both false belief and avoidance desire reasoning, participants are required to select from competing responses and inhibit the prepotent response (e.g., true versus believed location/desired versus undesired location). Consequently, false belief and avoidance desire states may draw on a domain-general 'selection processor', in order to direct executive selection resources in attentionally demanding situations. Importantly, avoidance desire (i.e. "desire to avoid") can concern objects or situations that are either intrinsically desirable or undesirable from the participant's own point of view. Hence, variation in the valence of desire does not reduce to a question of whether the participant shares the character's desire: indeed desire valence and self-other congruence of desire are logically orthogonal factors. To explain these findings, Leslie and colleagues suggest that EC has a more general role in ToM that is not restricted only to cases that require inhibition of self-perspective. Previous neuroimaging studies examining EC more generally in ToM have typically used separate tasks to identify ToM and EC regions. These indicate some overlap between neural regions recruited for EC tasks and false belief reasoning, extending beyond IFG to include anterior cingulate cortex (ACC), frontal operculum (FO) and frontal eyefields (FEF) (Rothmayr et al., 2011; Saxe et al., 2006b; van der Meer et al., 2011).

Whilst interest in the neural basis of executive function in ToM is growing, most previous studies are limited in their ability to cast light on the role of executive function in ToM. Most have sought to identify neural regions involved *only* in ToM by comparing activation observed in a ToM condition with that in a non-ToM control condition. Such approaches may enable powerful tests of hypotheses about brain regions that are domain-specific for ToM, but run the risk of subtracting out activation that is critical for understanding how ToM is achieved in the brain. A fruitful alternative approach is to manipulate psychologically relevant factors within a ToM task (for a discussion of these issues see discussion between Saxe et al., 2006a and Friston and Henson, 2006). Surprisingly few previous studies of ToM, however, have attempted such manipulations. Sommer et al. (2007) provide one of the few direct comparisons between true and false belief reasoning. In their nonverbal task, participants viewed a series of cartoons which depicted a true or false belief scenario analogous to the object transfer task outlined earlier. Regions which were more responsive to false belief over true belief attribution included the right TPJ, ACC and right lPFC. The reverse contrast only identified the superior frontal gyrus, which is in contention with the view that TPJ is an essential component when attributing any transient mental state (see van Overwalle, 2009). These data indicate that false belief reasoning might recruit EC regions. However, they are difficult to interpret with confidence, because it is not clear whether participants were solving the contrasting true belief condition by mental state ascription or by simply referring to the true state of affairs (Aichorn et al., 2009). Consequently, further examination of these two mental states is warranted, where attending to a protagonist's mental state is made unavoidable in both true and false belief reasoning. This was the case in the current study.

The neural basis of conative states such as desires has been studied less extensively. Hooker et al. (2008) examined neural activation when making empathic judgements for characters with varying perspectives. More directly relevant to the current study, Abraham et al. (2010) had participants read a series of short vignettes which varied the valence of belief and desire: either an agent's belief turned out to

[1] These variations in belief and desire both vary the difficulty of the belief–desire reasoning task. However, beliefs in the current study varied in terms of their consistency with the participant's self-perspective (true beliefs versus false beliefs), whereas desires varied only in terms of whether the target character liked or disliked the food. Therefore we use the term "valence" to refer collectively to these variations, so that true beliefs and desires for foods are described as "positively valenced" and false beliefs and desires to avoid foods are described as "negatively valenced".

be true or false, or an agent's desire turned out to be fulfilled or unfulfilled. The vignettes were followed by a yes/no question in which participants judged how the agent would feel about the true state of affairs. Their results were broadly consistent with the existing literature and showed recruitment of key mentalising areas including TPJ and mPFC for both the belief and desire conditions compared to a non-ToM reasoning task. An analysis of the overall effect of valence (of both belief and desire) identified activation in mid-line structures, including mPFC and posterior cingulate cortex. This study is interesting because it attempts to separate the demands of belief and desire reasoning into different experimental conditions. However, this also leads to limitations. Firstly, it is unclear whether this separation can be entirely successful, since judging an agent's feelings on the basis of his belief may lead participants to think about his desire even though they were not asked to. Likewise, judging an agent's feelings on the basis of her desire may lead participants to think about her belief. Secondly, it is unclear how such conditions relate to the canonical forms of ToM reasoning, in which we combine information about both belief and desire to predict or explain an agent's action. For this reason, the current study followed the longstanding literature on ToM in children by asking participants to predict a character's actions on the basis of his belief and desire.

We deployed a novel task (Apperly et al., 2011) based upon the object transfer action prediction ToM task, from which there are already considerable behavioural data (e.g., Friedman and Leslie, 2005; Wellman et al., 2001; Wimmer and Perner, 1983). Our previous work has shown this design to be able to detect differences in reaction time and error rate when participants predicted an agent's action on the basis of true versus false beliefs and a desire to approach versus avoid an object. This task allowed us to look specifically at neural activation during the decision making phase during which these behavioural effects are observed. The novel task comprises an orthogonal design whereby belief (true/false) and desire (approach/avoid) states are manipulated within a single, within-subjects experiment. The use of this factorial design enabled a whole brain analysis to isolate any neural regions that were modulated either by the valence of belief state, or by the valence of the desire state, or both. In doing so, the present study sought to address three key questions.

Do our factors of Belief-Valence and Desire-Valence recruit any regions of the ToM network?

It is entirely possible that our factors of Belief- and Desire-Valence would not recruit any regions of the ToM network, because beliefs and desires feature in all of our experimental conditions. It is important to emphasise that the present task and analyses were not designed to identify regions that are specifically involved in representing beliefs or desires in comparison with non-ToM reasoning, but instead were designed to identify those regions that are responsive to variation in the valence of either belief or desire during an action prediction. This is informative because, as reviewed above, previous work shows that the valence of beliefs and desires makes a critical contribution to the difficulty of belief–desire reasoning for both children and adults.

If our factors of Belief-Valence and Desire-Valence recruit regions of the ToM network, is this just because those regions are involved in attention/executive control, not because they are involved in ToM per se?

Although the literature converges on identifying brain regions that are consistently associated with ToM, the role of these regions remains controversial. On one view, at least some regions – in particular, some regions of right TPJ – are activated during ToM tasks because they are specifically involved in ToM (e.g., Saxe and Kanwisher, 2003; Scholz et al., 2009). On another view, such activation merely reflects the allocation or reorientation of attention, which is known to be a function of TPJ, and is a confounding feature of many ToM tasks (Mitchell, 2008; Rothmayr et al., 2011). The need for care on this question is emphasised by a recent structural imaging which demonstrated that TPJ can be subdivided in terms of its connectivity with other brain regions associated respectively with ToM and attention (Mars et al., 2011). To address this issue, we used a separate ToM localiser task (Saxe and Kanwisher, 2003) alongside our novel task. This localiser contrasts brain activation observed during false belief trials with that observed during closely-matched false photograph trials. It is widely agreed that false photograph tasks are an excellent match for most of the confounding demands that false belief tasks make on memory, EC and attention (e.g., Aichorn et al., 2009; Saxe and Powell, 2006), so although interpretation of this localiser remains controversial (e.g., Mitchell, 2008 vs Young et al., 2010), it is currently the best method available for identifying brain regions that might be specifically involved in ToM. By using the localiser alongside the belief–desire task, we were able to explore the neural signature of specific belief and desire states in those voxels within TPJ that appear to be specifically responsive to mental representation.

Do we observe differential activation of EC regions due to the Belief-Valence factor compared with the Desire-Valence factor?

The two theories of EC in ToM reviewed earlier make alternate, but not incompatible, predictions about the pattern of brain activation in belief and desire reasoning. Firstly, Leslie and colleagues' executive performance account of ToM (Friedman and Leslie, 2004, 2005; Leslie and Polizzi, 1998; Leslie et al., 2005) posits that both avoidance desire and false belief reasoning recruit common executive resources for the selective control of attention. If this theory is correct, negatively valenced belief and desire states will draw on the same executive regions. Secondly, if EC is involved in self-perspective inhibition (McCleery et al., 2011; Samson et al., 2005; van der Meer et al., 2011), then we should expect to see different recruitment of brain areas for the factors of Belief-Valence and Desire valence. This is because false belief trials are thought to make higher demands on self-perspective inhibition than true belief trials, whereas there is no systematic variation in the need for self-perspective inhibition when the agent has a desire to avoid rather than approach the object.

Method

Participants

Twenty healthy adults participated in both of the fMRI experiments. All gave informed ethical consent and were given course credit or a small honorarium for their participation. The study had appropriate research ethics approval from the University of Birmingham. One participant was excluded from all analyses due to poor behavioural task performance during scanning. The remaining 19 participants were included in all analyses (6 male, 13 female; age range 18–39, $\overline{X}$ age = 25 years). All participants were strongly right handed, measured with a modified form of the Annett Handedness Questionnaire (1970), and were proficient English speakers.

Materials and procedure

Pre-screen

Suitability to participate was determined several days prior to collecting the neuroimaging data. The Wide Range Achievement Test — Third Edition (WRAT-3) Reading Scale was administered to screen for reading disabilities and ensure reading proficiency commensurate with the experimental tasks. The participants then completed a computer based interactive training session which gave an overview of the belief–desire reasoning experiment. They then attempted one block of experimental trials outside of the MRI scanner. Only participants who performed above chance on this pre-test block took part in the fMRI experiments.

fMRI belief–desire reasoning experiment

The main experiment was based on a paradigm devised by Apperly et al. (2011) which was revised for use as an event-related design within the MRI scanner. Pilot work using the revised paradigm confirmed that experimental timings were appropriate in that the participants were able to perform the task to a high degree of accuracy (>90% correct trials). The experiment utilised an orthogonal design which had four equally occurring conditions that were based on a protagonist's belief state (true (B+) or false (B−)) and desire state (approach (D+) or avoid (D−)). By varying the protagonist's beliefs and desires four conditions were created: B+D+, B+D−, B−D+ and B−D−. Note that immediately prior to participating in the main experiment all participants completed one further practice block outside of the MRI scanner so as to refamiliarise themselves with the main experimental task. None of the pre-test or practice trials were used in the main fMRI experiment.

The experiment required participants to predict which one of two different coloured boxes a character would open based on a scenario in which the character would seek out food they love and avoid food they hate (Fig. 1).

A male protagonist, introduced during the training and practice sessions as Simon, was always used with male participants, whereas a female character, Sally, was always used with female participants. Each scenario consisted of three centre justified statements followed by a picture response probe then rest. Statements were separated by a fixation period of 400 ms. A variable interstimulus interval was used (range = 9000–14,000 ms, $\overline{X} = 11{,}500$ ms), during which a small fixation dot was displayed. The temporal order of the statement types was randomised, but all scenarios contained one belief statement (e.g., he thinks the chips are in the red box), one desire statement (e.g., he loves chips) and one reality statement (e.g., the chips are in the blue box). This design meant that participants were always explicitly told the character's belief, whether the belief was true or false. Moreover, randomisation of the statement order ensured that participants needed to encode the character's true belief on at least the 50% of trials on which they did not already know the object's true location. In these ways our design addressed the weakness of earlier studies in which participants could safely ignore a character's beliefs on true belief trials, relying instead on their own knowledge of reality.

The statements were followed by a response probe. If the protagonist appeared in the response probe, participants indicated whether the character would open the left or the right box based on the agent's belief–desire state, using a two button box placed in their left hand. These were the trials of interest and made up two thirds of the overall number of trials presented. In the other one third of trials, the protagonist was replaced with a question mark in the response probe. In this instance, participants responded by giving the true location of the food. These anti strategy trials were used to ensure that the participants had to attend to all three statements, and did not form any part of the analyses presented within the present paper. Twelve different food types were used, which were consistently “loved” or “hated” by the on screen protagonist. Food preferences were counterbalanced so that half of the participants saw one consistent set of preferences, whereas the other half saw the opposite preferences. The correct response corresponded to the left and right box an equal number of times. Participants completed four blocks of trials, each of which contained 24 trials (16 trials of interest, 8 anti strategy trials). Each block lasted 7 min 22 s which included an initial instruction and final thank you screen.

fMRI theory of mind localiser experiment

The localiser task was substantially based on the experimental procedure devised by Saxe and Kanwisher (2003). Stimuli consisted of a subgroup of the current localiser stories (see Saxe and Andrews-Hanna, n.d.), some of which were anglicised for use in the present experiment. Participants read a total of 24 short vignettes which referred to either a protagonist's false belief (FB) or an outdated physical representation, such as the false photograph scenario (FP). Each vignette was displayed for 10 s, which was followed for 4 s with a short true or false question about the preceding story. This required participants to make a response using a two button box that was placed in their left hand. Stories alternated between FB and FP and were interleaved with a

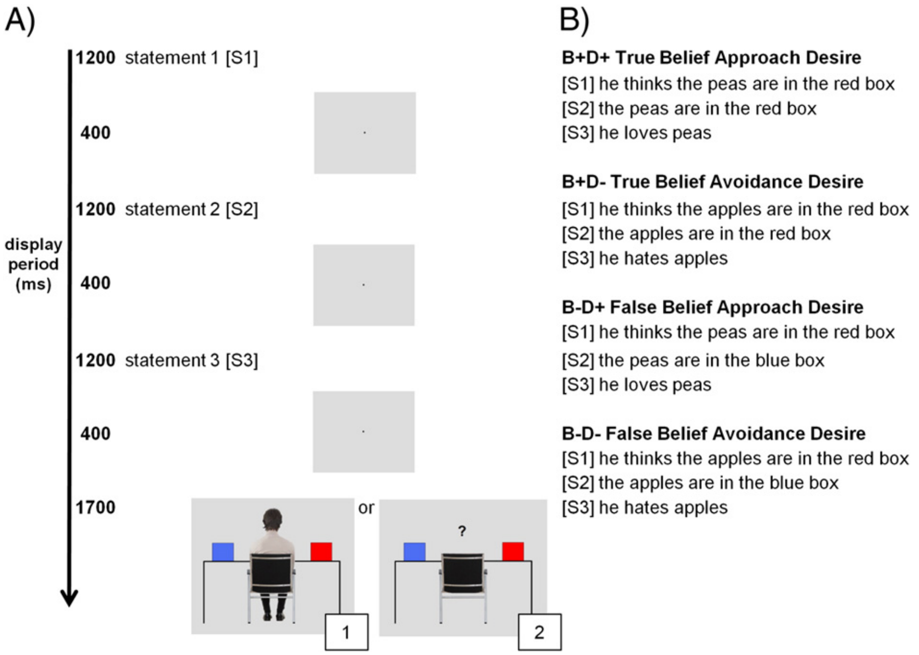
A)
1200 statement 1 [S1]
400
1200 statement 2 [S2]
display period (ms)
400
1200 statement 3 [S3]
400
1700
or
?
1
2
B)
B+D+ True Belief Approach Desire
[S1] he thinks the peas are in the red box
[S2] the peas are in the red box
[S3] he loves peas
B+D- True Belief Avoidance Desire
[S1] he thinks the apples are in the red box
[S2] the apples are in the red box
[S3] he hates apples
B-D+ False Belief Approach Desire
[S1] he thinks the peas are in the red box
[S2] the peas are in the blue box
[S3] he loves peas
B-D- False Belief Avoidance Desire
[S1] he thinks the apples are in the red box
[S2] the apples are in the blue box
[S3] he hates apples

Fig. 1. (Panel A) Experimental sequence of a single trial. Response probe 1 is an example of the image displayed for a trial of interest, picture 2 is an example of the response probe displayed during the anti-strategy trials. Note that the white numbered boxes were not part of the stimuli. (Panel B) Example trial sequences for each of the four conditions. The order of statement types was randomised for each trial.

13.5 s rest period. The localiser experiment comprised of four blocks of six trials, each containing three of each type of story. Participants were given four practice trials immediately prior to scanning to familiarise themselves with the localiser task.

Data acquisition and analysis

Neuroimaging data acquisition and processing

Each participant's data were acquired during a single scanning session using a 3 T Philips Achieva scanner. All stimuli were presented using Presentation software (Neurobehavioural Systems, CA) which also recorded the behavioural response data simultaneously. Participants completed two blocks of the main belief–desire experiment followed by all four blocks of the localiser task and the remaining two blocks of the main experiment. 177 T2*-weighted echo-planar imaging (EPI) volumes were obtained per block of the belief–desire experiment and 77 EPI volumes were acquired for each block of the localiser task. Both tasks utilised the same general imaging parameters to achieve whole brain coverage (TR = 2.5 s, TE = 35 ms, acquisition matrix = 96 × 96, flip angle = 83°, voxel size = 3 × 3 × 3 mm^3). EPI images consisted of 44 axial slices that were obtained consecutively in a bottom up sequence. High resolution T1-weighted structural images were acquired following collection of the functional data (1 × 1 × 1 mm^3 isotropic voxels).

Preprocessing and statistical analyses of the data were performed using the FMRIB software library (FSL version v.5.98; FMRIB, Oxford, www.fmrib.ox.ac.uk/fsl). For both experiments, initial preprocessing of the functional data consisted of slice timing correction, and motion correction using rigid body transformations (MCFLIRT). The blood oxygen level dependent (BOLD) signals were high-pass filtered using a Gaussian weighted filter of 30 s for the belief–desire task and 21 s for the localiser task. The BOLD data were then spatially smoothed using a 5 mm full-width-half-maximum kernel. The functional data were registered to their respective structural images and transformed to a standard template based on the Montreal Neurological Institute (MNI) reference brain, using a 6-DoF linear transformation (FLIRT).

Belief–desire reasoning experiment analysis

The functional data resulting from the four conditions were modelled as four explanatory variables (EVs) of interest: B + D+, B + D−, B − D+, B − D−. To focus on the decision making phase of the sequence, the onset of each event was time locked to when the participant made a button response for each trial. Each EV comprised an arbitrary duration of 100 ms. The EVs were convolved with a gamma derived haemodynamic response function (HRF) within a general linear model framework (GLM). Motion parameters were treated as regressors of no interest in order to account for unwanted motion effects. The sentence phase was modelled as a regressor of no interest and orthogonalised with respect to the main EVs. Session data were aggregated per participant using a second level fixed effects model. These 19 second level models were used to provide the input data for ROI analyses. Third level modelling was used to aggregate the data across participants in a 2 × 2 repeated measures ANOVA with Belief-Valence (B+/B−) and Desire-Valence (D+/D−) as within subjects factors. The final whole brain result was based on a mixed effects (ME) analysis with cluster based thresholding at $Z > 2.3$, $p_{corr} < 0.05$.

ToM localiser experiment analysis

The localiser task was modelled as per Saxe and Kanwisher (2003). Statistical analysis was conducted using a GLM. Two EVs which reflected the two conditions, FB and FP, were convolved with a gamma-derived HRF. Second and third level modelling was used to aggregate the data across sessions and participants for the contrast of interest FB > FP. For examination of activation between the two experimental paradigms, post-stats processing of the group result was conducted as per the parameters used for the main belief–desire reasoning task (ME analysis, $Z > 2.3$, $p_{corr} < 0.05$).

Overlap analysis

Using the whole brain data, any overlap between activations from the localiser task and the belief–desire task were identified using FSL's command line tools (fslmaths). A logical AND function was applied to the thresholded data ($p_{corr} < 0.05$, $Z > 2.3$) for the factors of Belief-Valence and Desire-Valence and the localiser FB > FP contrast.

ROI analysis

ROI masks were created using the MarsBaR region of interest toolbox (version 0.42 marsbar.sourceforge.net) for SPM8 (www.fil.ion.ucl.ac.uk/spm). Masks comprised a sphere with a 5-mm radius centred on the single subject peak voxel within TPJ for the FB > FP localiser contrast. ROI analyses were carried out on each participant's aggregated sessional data for the 4 EVs modelled in the main belief–desire experiment. The mean percentage signal change (PSC) for each condition of interest within each ROI was extracted using FSL Featquery (www.fmrib.ox.ac.uk/fsl/feat5/featquery.html).

Results

Belief–desire reasoning task behavioural results

All reaction times (RTs) were recorded from the onset of the response probe. Any incorrect responses or data points that were 2 standard deviations outside of the participant's condition mean were removed for RT analysis. A 2 × 2 repeated measures ANOVA was conducted on the remaining data, with Belief-Valence (B+/B−) and Desire-Valence (D+/D−) as within subjects factors. This revealed significant main effects of Belief-Valence, where B− > B+ ($F(1,18) = 46.94$, $p < 0.001$, $\eta^2 = 0.72$) and Desire-Valence, where D− > D+ ($F(1,18) = 25.21$, $p < 0.001$, $\eta^2 = 0.58$) but no interaction ($F(1,18) = 0.21$, $p = 0.66$, $\eta^2 = 0.01$). Fig. 2A summarises the mean RT for correct responses given across the four conditions.

The participant's error rate was analysed in a further 2 × 2 repeated measures ANOVA. This also indicated significant main effects of Belief-Valence where B− > B+ ($F(1,18) = 22.55$, $p < 0.001$, $\eta^2 = 0.56$) and Desire-Valence where D− > D+ ($F(1,18) = 5.86$, $p = 0.03$, $\eta^2 = 0.25$), but no interaction between the two ($F(1,18) = 0.63$, $p = 0.44$, $\eta^2 = 0.03$). Fig. 2B illustrates the mean proportion of incorrect responses.

Whole brain analysis

Belief–desire reasoning task

A 2 × 2 repeated measures ANOVA of the belief–desire reasoning task identified main effects of Belief-Valence (B+/B−) and Desire-Valence (D+/D−) but no interaction between the two factors. Manipulation of Belief-Valence recruited bilateral TPJ, superior parietal and occipital cortices, as well as frontal areas including the ACC (BA 32), bilateral dorsolateral prefrontal cortex (dlPFC) (BA 9, 46) and vlPFC including bilateral orbital frontal cortex, IFG and FO (BA 44, 45, 47) (Table 1; red shading in Fig. 3). Varying Desire-Valence also elicited activation in bilateral TPJ, superior parietal and occipital cortices, and medial frontal regions including the ACC. However, in contrast to the factor of Belief-Valence, frontal activation was largely left lateralised, spanning both dlPFC and superior regions of vlPFC. Modulation of right frontal areas was limited to dlPFC (Table 2; green shading in Fig. 3). Thus, whilst the valence of belief and desire both modulated activation in ACC, only belief was shown to influence the most inferior parts of vlPFC.

Localiser task

A mixed effects analysis of the whole brain localiser data identified neural regions that were more responsive to mental than physical

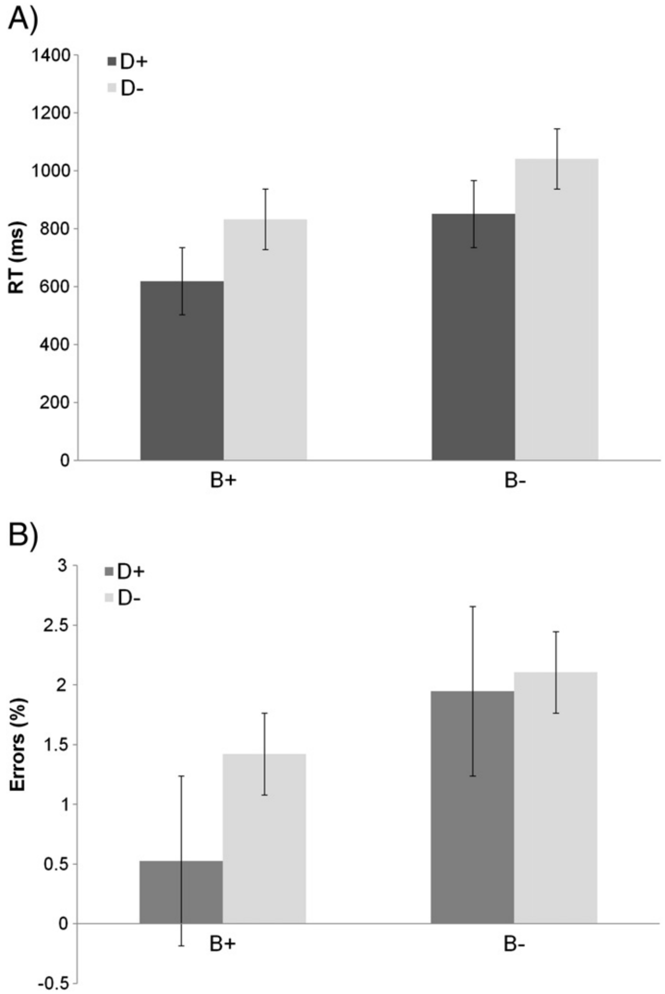

Fig. 2. Error bars reflect +/− 1 SE of the mean. (Panel A). Group mean reaction time per condition for correct responses (ms): B+D+ = 619.00; B+D− = 832.39; B−D+ = 851.12; B−D− = 1041.16. (Panel B). Percentage of errors made per condition: B+D+ = 0.17%; B+D− = 0.46%; B−D+ = 0.67%; B−D− = 0.86%.

representation (FB > FP, $p_{corr}<0.05$). These results were consistent with previous ToM studies, showing that the FB > FP contrast recruits core regions of the ToM network, including bilateral TPJ and mPFC (Table 3; green shading in Fig. 4A).

Overlap analysis results

Inspection of the activation maps from the group data suggested considerable overlap between neural regions recruited by the localiser task and the belief–desire reasoning task, as shown in Fig. 4A. An overlap analysis identified that only bilateral TPJ was required for all three variations of mentalising (Fig. 4B).

ROI results

As bilateral TPJ were the only regions identified for both mental representation (localiser) and variation in mental state valence (belief–desire task), we focused ROI analyses on these areas. ROIs were identified using the localiser task in 18 of 19 individual participants in the rTPJ and 18/19 in lTPJ. ROI analysis was conducted on data from the belief–desire reasoning task and a 2 × 2 repeated measures ANOVA conducted on the mean PSC data for each ROI (Fig. 5). The right TPJ's response was higher when reasoning about a false than a true belief ($F(1,17)=20.43$, $p<0.01$, $\eta^2=0.55$) and higher for avoidance versus approach desire ($F(1,17)=9.47$, $p<0.01$, $\eta^2=0.36$). No interaction existed ($F(1,17)=2.87$, $p=0.11$, $\eta^2=0.14$). Similar effects were detected in lTPJ where its response was higher when reasoning about a false- than a true-belief ($F(1,17)=12.73$, $p<0.01$, $\eta2=0.43$) and higher for negative- versus positive-desire ($F(1,17)=28.64$, $p<0.001$, $\eta2=0.63$), but no interaction existed between Belief- and Desire-Valence ($F(1,17)=1.97$, $p=0.18$, $\eta2=0.10$).

Discussion

Behavioural evidence suggests that negatively valenced mental states – false beliefs and avoidance desires – are more difficult to process than their positively valenced counterparts. On developmentally sensitive tasks, young children pass false belief and avoidance desire tasks at a later age than true belief and approach desire tasks (Cassidy, 1998). Suitably adapted tasks demonstrate that adult participants, too, show a similar pattern of relative difficulty, reflected in response times and residual error rates (Apperly et al., 2011; German and Hehman, 2006). Moreover, in both children and adults, performance on such tasks is associated with independent tests of EC (e.g., Carlson and Moses, 2001; German and Hehman, 2006). The neuroimaging literature consistently identifies TPJ and mPFC as core ToM regions, but less is known about how activity in these regions is modulated by psychologically relevant differences between positive and negative valencies. Likewise, little is known about how and when neuro-cognitive systems for EC are recruited in the service of different aspects of ToM. We addressed these issues in the current study by manipulating the valence of belief and desire states and by examining neural activity during the response phase of each trial, during which the behavioural costs of belief–desire reasoning have been observed on this task.

Do our factors of Belief-Valence and Desire-Valence recruit any regions of the ToM network?

We set out to investigate how variation in the valence of belief and desire states affects recruitment of the ToM network. A whole brain

Table 1
Cluster peaks for the belief–desire reasoning task: factor of Belief-Valence.

Hemisphere and region	Brodmann areas	Cluster size (voxels)	Peak MNI coordinates			Z-values
			x	y	z	
L inferior frontal gyrus, L middle frontal gyrus, L frontal operculum, L frontal orbital cortex	6, 8, 9, 38, 44, 45, 46, 47, 48	3134	−50	20	24	4.97
L temporoparietal junction, L supramarginal gyrus, L lateral occipital cortex	22, 39, 40	2859	−54	−52	26	4.92
R orbital frontal cortex, R frontal operculum, R inferior frontal gyrus, R middle frontal gyrus	9, 38, 44, 45, 46, 47, 48	1414	34	24	−6	4.85
R temporoparietal junction, R lateral occipital cortex, R middle temporal gyrus	22, 39, 40	1411	52	−54	24	4.78
L/R superior frontal gyrus, L/R paracingulate gyrus, L/R anterior cingulate cortex	8, 9, 24, 32	2069	0	28	46	4.37
R cerebellum crus I	–	561	18	−70	−34	3.78

Note. Clusters reflect results of 2-way repeated measures ANOVA for the factor of belief-valence (B+/B−). The table shows neural regions which are modulated by varying truth-status (true/false), $p_{corr}<0.05$.

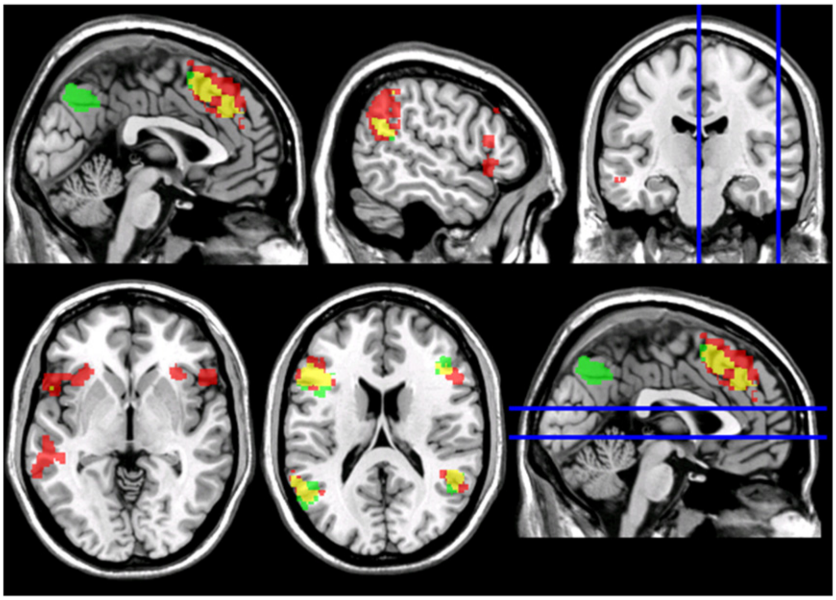

Fig. 3. Result from 2×2 repeated measures ANOVA whole brain analysis of the belief–desire reasoning task, with Belief-Valence (B+/B−) and Desire−Valence (D+/D−) as within-subjects factors. Selected slices highlight modulation in ToM and EF regions for the factors of Belief-Valence (red) and Desire-Valence (green). Yellow areas indicate regions recruited by both factors (B/D). The group data are overlaid on the MNI brain template, showing significantly activated voxels where $Z > 2.3$, $p_{corr} < 0.05$. Slices from top left to bottom right, x = −1, 54; z = −2, 18 respectively. Images reflect Z-corrected *F*-stat images and are displayed in neurological convention, where left is represented on the left side of the image

analysis demonstrated that variation in mental state valence modulates activity in neural regions regularly implicated in general ToM tasks including temporoparietal, medial parietal and some prefrontal regions. This finding converges with evidence from a small number of studies that suggest that these regions not only respond to ToM tasks in contrast to non-ToM baseline tasks, but also that their activity varies according to the valence of belief and/or desire (Abraham et al., 2010; Sommer et al., 2007; van der Meer et al., 2011). Importantly, we find these effects during a canonical ToM task that requires participants to predict the action of an agent on the basis of belief and desire.

Alongside TPJ, anterior rostral areas of the mPFC are also commonly implicated in studies of ToM (Amodio and Frith, 2006; Carrington and Bailey, 2009; Lieberman, 2007; Mar, 2011; van Overwalle, 2009). Whilst belief and desire reasoning modulated dorsal areas of the medial frontal cortex – particularly dorsal ACC – our novel paradigm showed no activation in anterior rostral mPFC. This finding contrasts with the ToM localiser task, which did show activity in anterior rostral mPFC. We believe that this pattern may be understood on the hypothesis that rostral mPFC is recruited for ToM to the degree that participants must go beyond the information immediately available to them, making social inferences about traits and norms, engaging in self-reflection or episodic thinking about the past or future (Gilbert et al., 2006). Such requirements are common in laboratory tasks and in everyday ToM, but are not a necessary feature of ToM cognition. In our belief–desire task participants were directly informed of the character's mental states, and the correct prediction of his or her action was wholly determined by deductive reasoning from this information. Thus, although participants needed to represent and reason about mental states, there was simply no need for inferences about traits, self reflection or episodic thinking. In contrast, the localiser task involved vignettes that, though short, did require participants to construct a situational context in which the character's mental states might be inferred. We suggest that it may be this need for elaborative processing that results in the recruitment of rostral mPFC in the service of ToM inferences.

If our factors of Belief-Valence and Desire-Valence recruit regions of the ToM network, is this just because those regions are involved in attention/executive control, not because they are involved in ToM per se?

There are three main alternate explanations as to the role of TPJ in ToM. One is that this region responds specifically to transient mental states, regardless of their content or status; thus, TPJ may be specialised towards ToM (e.g., Saxe and Kanwisher, 2003; van Overwalle, 2009). Support for this theory is found in data which pinpoint TPJ for a variety of ToM, but not control, tasks. This includes the attribution of beliefs (e.g., Aichorn et al., 2009; Saxe and Kanwisher, 2003; Saxe and Wexler, 2005; Scholz et al., 2009) and, although little

Table 2
Cluster peaks for the belief–desire reasoning task: factor of Desire-Valence.

Hemisphere and region	Brodmann areas	Cluster size (voxels)	Peak MNI coordinates			Z-values
			x	y	z	
L middle frontal gyrus, L inferior frontal gyrus	6, 9, 44, 45, 48	2339	−46	12	36	5.37
L/R precuneus	7	736	2	−66	42	4.63
L angular gyrus, L temporoparietal junction, L lateral occipital cortex	21, 39, 40	2141	−40	−56	52	4.46
L/R superior frontal gyrus, L/R paracingulate gyrus, L/R anterior cingulate cortex	8, 9, 24, 32	767	2	18	54	4.33
R angular gyrus, R temporoparietal junction, R supramarginal gyrus	7, 22, 40, 41, 48	1067	36	−48	42	3.8
R inferior frontal gyrus, R middle frontal gyrus	9, 44, 45, 48	467	44	28	20	3.69

Note. Clusters reflect results of 2-way repeated measures ANOVA for the factor of desire-valence (D+/D−). The table shows neural regions which are modulated by varying desire-status (approach/avoid), $p_{corr} < 0.05$.

Table 3
Cluster peaks for the ToM localiser task, showing activation where FB > FP.

Hemisphere and region	Brodmann areas	Cluster size (voxels)	Peak MNI coordinates			Z-values
			x	y	z	
R temporoparietal junction, R lateral occipital cortex, R middle temporal gyrus	21, 22, 39, 40, 42	4157	60	−58	18	4.76
L/R precuneus	7	2866	2	−58	36	5.56
L/R frontal pole, L/R medial prefrontal cortex, L/R superior frontal gyrus	8, 9, 10, 11	3511	−4	66	−12	4.65
L middle temporal gyrus	20, 21	1441	−58	−8	−20	4.72
L lateral occipital cortex, L temporoparietal junction, L angular gyrus	7, 19, 21, 39	1263	−42	−70	−38	4.08
L cerebellum crus II	–	957	−30	−80	−40	3.84
L cerebellum IX	–	382	−4	−56	−46	3.92

Note. Clusters reflect results from *t*-test of FB > FP. Table shows neural regions which are more responsive to false-belief than false-photo stimuli, $p_{corr} < 0.05$.

explored, desires (Saxe and Kanwisher, 2003). The second possibility is that TPJ may regulate the distinction between self and other (e.g. see Brass et al., 2009; Decety and Lamm, 2007). Activation of TPJ is a consistent feature of both mentalising and seemingly disparate tasks such as the inhibition of imitative behaviour. It has therefore been suggested that TPJ is recruited for situations which require a person to disengage self from other, so that an individual can appropriately assign behaviours or mental states as belonging to an external agent. Lastly, it has been suggested that TPJ activation is observed in ToM tasks because TPJ supports domain-general processes that are unintended confounds of ToM tasks, such as reorienting spatial attention away from miscued locations (e.g., Mitchell, 2008; Rothmayr et al., 2011). When applied, for example, to a false belief scenario, this process might reflect the need to redirect one's attention from location A ("true" location) to location B ("false" location). It is suggested that, as a result, ToM and exogenous attention tasks mutually activate right TPJ, which indicates that there may be some shared attentional component between ToM and spatial reorienting (Mitchell, 2008; Rothmayr et al., 2011).

Our findings do not fit well with the last of these three possibilities. The localiser task subtracted activation observed during false belief trials (which involve reasoning about false beliefs and management of attention between "false" and "true" locations) from activation observed during false photograph trials (which involve reasoning about photographs that are outdated/false and management of attention between "false" and "true" locations). Since the need to manage attention between "false" and "true" locations is present in both the false belief and false photograph conditions, and indeed, appears present to a similar degree, little activation due to such attention management is likely to survive the subtraction between these conditions. Instead, the surviving activation is more likely to be due to a difference between reasoning about false beliefs compared with false photographs. It is noteworthy, then, that this surviving activation in bilateral TPJ overlaps substantially with regions modulated by our novel belief–desire task. We think it unlikely that the common activation across these comparisons is due to a confounding requirement to reorient attention that has nothing to do with ToM.

Our findings also pose a challenge for the claim that TPJ is specialised for ToM and responds specifically to such transient mental states, regardless of their content or status (e.g., Saxe and Kanwisher, 2003; van Overwalle, 2009), because we found that activity in these regions was modulated by the valence of both beliefs and desires. However, our findings might be reconciled with this theoretical interpretation by supposing that TPJ is playing a similar functional role across these conditions, but its activity is up- or down-regulated by the relative difficulty of the different belief–desire conditions. The participants in the present study were slower to respond to both false belief and avoidance desire scenarios, and for this reason alone, activity in TPJ may have been held high for longer, or held higher overall. A further possibility is that TPJ is playing distinct functional roles across our belief and desire conditions, due to differential demands of representing true versus false beliefs and approach versus avoidance desires, or of making action predictions on the basis of this information. One potential source of differential demands is the need to maintain a distinction between self and other (e.g. see Brass et al., 2009; Decety and Lamm, 2007), though this need varies much more obviously between true and false beliefs than between positive and negative desires. What is potentially interesting in this general interpretation is that it offers a way of combining the insights of the other two: on the one hand TPJ recruitment during ToM tasks may not be due to confounding demands on attentional control

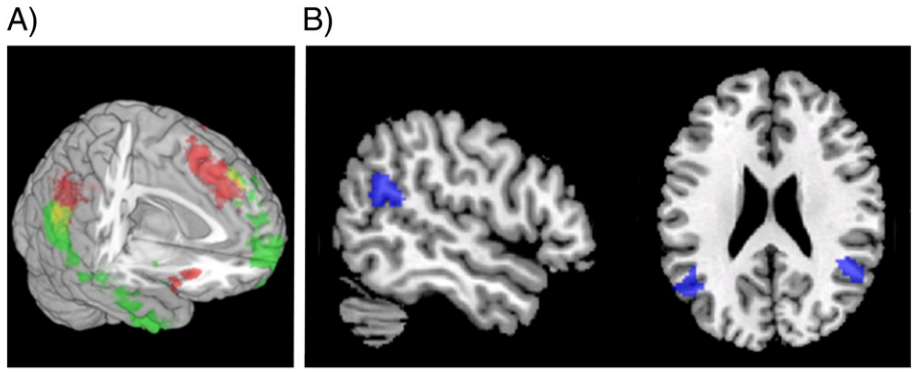

Fig. 4. (Panel A) Activation map for the contrast FB > FP (green) shown with the cluster maps from the belief–desire reasoning task, where the factors of Belief-Valence and Desire-Valence are represented by a single colour (red). Yellow areas indicate regions recruited for both the localiser and the belief–desire reasoning tasks. Each map is overlaid onto the MNI brain template and shows significantly activated voxels where $Z > 2.3$, $p_{corr} < 0.05$. (Panel B) Blue clusters reflect conjunction between localiser contrast FB > FP, and Belief-Valence Desire-Valence factors B+/B− and D+/D−, $p_{corr} < 0.05$. Slices $x = 52$, $z = 24$. Images reflect Z-corrected t-stat images and are displayed in neurological convention, where left is represented on the left side of the image.

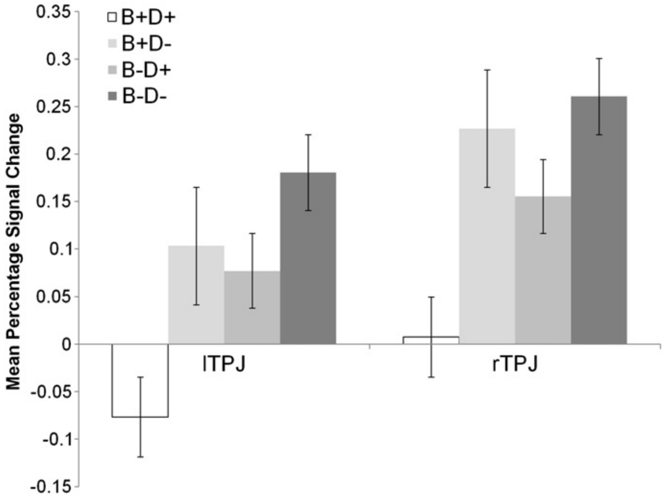

Fig. 5. Error bars reflect +/− 1 SE of the mean. Results from the ROI analysis, where ROI masks generated using the localiser task were applied to the belief–desire task. Group mean percentage signal change (PSC) per condition lTPJ: B + D + = −0.07; B + D − = 0.10; B − D + = 0.08; B − D − = 0.18. rTPJ: B + D + = 0.01; B + D − = 0.23; B − D + = 0.16; B − D − = 0.26.

in ToM tasks, on the other it may be that attentional control is intrinsic to ToM problems, not least in order to maintain and switch between representations of self and other.

Do we observe differential activation of EC regions due to the Belief-Valence factor compared with the Desire-Valence factor?

Leslie and colleagues find that false belief and avoidance desire will attract greater processing costs than true belief and avoidance desire (Friedman and Leslie, 2004, 2005; Leslie and Polizzi, 1998; Leslie et al., 2005). Our data converge with these findings and the wider literature on behavioural performance in adult belief–desire reasoning (Apperly et al., 2011; German and Hehman, 2006). Leslie and colleagues additionally specify that belief and desire reasoning is supported by a common process (termed a 'selection processor' in their account) which directs executive selection resources in attentionally demanding situations, for example, when attributing negatively valenced mental states. Our data are consistent with this idea and identify ACC as a possible candidate for EC processes associated with such variation in task difficulty. Whilst most extensively examined in the cognitive literature (e.g., Botvinick et al., 1999, 2004; Carter et al., 1998), ACC is increasingly acknowledged to play an important role in supporting social cognition (Amodio and Frith, 2006; Lieberman, 2007). Converging electrophysiological and neuroimaging data suggest a functional division within ACC, where dorsal areas subserve conflict monitoring and error detection, and rostral–ventral areas are primarily involved in the assessment of motivational or emotional information (Amodio and Frith, 2006; Bush et al., 2000; Devinsky et al., 1995). For the present experiment, both the valence of both belief and desire states was shown to modulate activation in dorsal ACC, suggesting that reasoning about very basic belief and desire states draws on a common cognitive process. As seen in our behavioural data, manipulation of mental state valence yielded processing costs in terms of error rates and response latencies. On this basis we propose that dorsal ACC indexes conflict (between self and other perspectives, and between the agent's belief about the object and his desire to avoid it) in order that further executive processes, such as inhibition and selection, may be initiated.

We have suggested that increased attentional demands may help explain behavioural difficulty with negatively valenced mental states, but it may be that this does not exhaust the role of EC in ToM. As described in the Introduction, a growing body of research suggests that participants will be slower and more error prone when holding in mind mental states which are incongruent with their own self perspective, such as a false belief or conflicting (not merely avoidance) desire state (Ruby and Decety, 2003; Samson et al., 2005; van der Meer et al., 2011). In the present study, we manipulated congruence with self other perspectives by asking participants to make predictions about a protagonist's behaviour in true and false belief scenarios. In contrast, our manipulation of approach versus avoidance desire did not result in differences in congruence of self and other perspectives, and so did not vary the need for self-perspective inhibition.

Belief-Valence, but not Desire-Valence, was seen to recruit the most inferior parts of bilateral vlPFC. Variation in the conflict between the perspectives of the participant and of the agent was manipulated in the Belief-Valence, but not Desire-Valence, condition. Thus, our data are consistent with the view that activation in vlPFC is modulated by variation in the need for self perspective inhibition, and show that this is a critical difference between true and false belief trials, as well as between false belief trials between which the salience of self-perspective is experimentally varied (Samson et al., 2005; van der Meer et al., 2011). The present dataset therefore provides strong evidence for a distinct role for EC beyond the generic control of attention during ToM tasks. In addition, EC is necessary when a perspective difference between self and other exists, as is the case for false belief. This converges with behavioural data from the current study and others indicating that knowledge of the true state of affairs interferes with the ability to select the believed (i.e. false) location, when the real and believed locations are incongruent, giving rise to the well-known phenomenon of egocentric biases and errors (Bernstein et al., 2004; Birch and Bloom, 2004, 2007; Mitchell and Lacohee, 1991). The process of inhibiting this self perspective, we suggest, specifically recruits vlPFC. Importantly, such activity would necessarily be missed in studies using the best-controlled comparisons between ToM and non-ToM tasks. For example, it would not be observed in Saxe and Kanwisher, 2003 ToM localiser because both the false belief and the false photograph conditions require inhibition of self perspective, and so any associated activation would be lost in the subtraction of one condition from another.

Conclusion

The present study provides evidence that converges with and extends a number of findings concerning the functional and neural basis of ToM. We find evidence that activation in TPJ is modulated by the valence of mental states, suggesting that this region is not responsive to transient mental states per se (e.g., see Saxe and Kanwisher, 2003; van Overwalle, 2009), but rather the content of such mental states. We find evidence that the mere requirement to represent a mental state may be insufficient to recruit rostral mPFC, but that this region is recruited when mental states need to be inferred on the basis of contextual information, consistent with Amodio and Frith (2006) and van Overwalle (2009). We also find evidence of the recruitment of neural regions associated with EC, which converges with behavioural evidence that ToM problems often require domain-general EC processes, as well as processes that might be more specific to ToM (Apperly et al., 2008, 2011; Carlson and Moses, 2001; Carlson et al., 1998, 2002; Cassidy, 1998; Friedman and Leslie, 2004, 2005; German and Hehman, 2006; Leslie and Polizzi, 1998; Leslie et al., 2005; Perner and Lang, 1999).

The present study significantly extends understanding of the relationship between ToM and EC, and the neural systems that support these abilities. ToM problems that participants find more difficult to solve – such as those involving false belief and avoidance desire – result in greater activity in neural systems involved in attentional control, such as ACC, and also in parts of the "ToM network", such as TPJ. Importantly, this effect of general difficulty can be distinguished from a more specific effect due to the need to resist interference from self perspective. This need only arises when there is a perspective difference between self and other – as in the false belief condition of the current study – and appears to recruit vlPFC in a distinctive manner. Nonetheless, additional work is required to further examine the role of EC in ToM and, in particular, the involvement of vlPFC in inhibition of self-perspective. The use of an established EC paradigm in parallel with a tightly controlled ToM task, such as was presented here, would advance our understanding of the neural basis of those domain-general processes that support ToM. Moreover, specific manipulations in terms of desire reasoning, where an agent's desire state is made systematically congruent or incongruent with self, would serve to further delineate the role of vlPFC in inhibition of self-perspective.

In sum, we demonstrate how the virtues of subtractive, "localiser" methods and methods that allow psychologically relevant parameters to be varied orthogonally may be combined to give a deeper understanding of the cognitive and neural basis of ToM than would be possible with either method alone.

Acknowledgments

This research was supported by the Economic and Social Research Council, award number ES/G01258X/1.

References

Abraham, A., Rakoczy, H., Wening, M., von Cramon, Y.D., Schubotz, R.L., 2010. Matching mind to world and vice versa: functional dissociations between belief and desire mental-state processing. Soc. Neurosci. 5, 1–18.

Aichorn, M., Perner, J., Weiss, B., Kronbichler, M., Staffen, W., Ladurner, G., 2009. Temporo-parietal junction activity in Theory-of-Mind tasks: falseness, beliefs, or attention. J. Cogn. Neurosci. 21 (6), 1179–1192.

Amodio, D.M., Frith, C.D., 2006. Meeting of minds: the medial frontal cortex and social cognition. Nat. Rev. Neurosci. 7, 268–277.

Apperly, I.A., Samson, D., Humphreys, G.W., 2005. Domain-specificity and theory of mind: evaluating evidence from neuropsychology. Trends Cogn. Sci. 9 (12), 572–577.

Apperly, I.A., Back, E., Samson, D., France, L., 2008. The cost of thinking about false beliefs: evidence from adults' performance on a non-inferential theory of mind task. Cognition 106 (3), 1093–1108.

Apperly, I.A., Samson, D., Humphreys, G.W., 2009. Studies of adults can inform accounts of theory of mind development. Dev. Psychol. 45 (1), 190–201.

Apperly, I.A., Warren, F., Andrews, B.J., Grant, J., Todd, S., 2011. Developmental continuity in theory of mind: speed and accuracy of belief desire reasoning in children and adults. Child Dev. 82 (5), 1691–1703.

Bernstein, D.M., Atance, C., Loftus, G.R., Meltzoff, A., 2004. We saw it all along—visual hindsight bias in children and adults. Psychol. Sci. 15 (4), 264–267.

Birch, S.A.J., Bloom, P., 2004. Understanding children's and adults' limitations in mental state reasoning. Trends Cogn. Sci. 8, 255–260.

Birch, S.A.J., Bloom, P., 2007. The curse of knowledge in reasoning about false beliefs. Psychol. Sci. 18 (5), 382–386.

Botvinick, M., Nystrom, L.E., Fissell, K., Carter, C.S., Cohen, J.D., 1999. Conflict monitoring versus selection-for-action in anterior cingulate cortex. Nature 402, 179–181.

Botvinick, M.M., Cohen, J.D., Carter, C.S., 2004. Conflict monitoring and anterior cingulate cortex: an update. Trends Cogn. Sci. 8 (12), 539–546.

Brass, M., Ruby, P., Spengler, S., 2009. Inhibition of imitative behaviour and social cognition. Philos. Trans. R. Soc. Lond. B Biol. Sci. 364, 2359–2367.

Bush, G., Luu, P., Posner, I., 2000. Cognitive and emotional influences in anterior cingulate cortex. Trends Cogn. Sci. 4 (6), 215–222.

Carlson, S.M., Moses, L.J., 2001. Individual differences in inhibitory control and children's theory of mind. Child Dev. 72 (4), 1032–1053.

Carlson, S.M., Moses, L.J., Hix, H.R., 1998. The role of inhibitory processes in young children's difficulties with deception and false belief. Child Dev. 69, 672–691.

Carlson, S.M., Moses, L.J., Breton, C., 2002. How specific is the relation between executive function and theory of mind? Contributions of inhibitory control and working memory. Infant Child Dev. 11, 73–92.

Carrington, S.J., Bailey, A.J., 2009. Are there theory of mind regions in the brain? A review of the neuroimaging literature. Hum. Brain Mapp. 30 (8), 2313–2335.

Carter, C.S., Braver, T.S., Barch, D.M., Botvinick, D.N., Cohen, J.D., 1998. Anterior cingulate cortex, error detection, and the online monitoring of performance. Science 280, 747–749.

Cassidy, K.W., 1998. Three- and four-year-old children's ability to use desire- and belief-based reasoning. Cognition 66, B1–B11.

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 (6), 580–593.

Devinsky, O., Morrell, M.J., Vogt, B.A., 1995. Contributions of anterior cingulate cortex to behaviour. Brain 118, 279–306.

Friedman, O., Leslie, A.M., 2004. Mechanisms of belief desire reasoning: inhibition and bias. Psychol. Sci. 15, 547–552.

Friedman, O., Leslie, A.M., 2005. Processing demands in belief desire reasoning: inhibition or general difficulty? Dev. Sci. 8, 218–225.

Friston, K.J., Henson, R.N., 2006. Commentary on divide and conquer: a defense of functional localizers. Neuroimage 30, 1097–1099.

German, T.P., Hehman, J.A., 2006. Representational and executive selection resources in 'theory of mind': evidence from compromised belief desire reasoning in old age. Cognition 101, 129–152.

Gilbert, S.J., Spengler, S., Simons, J.S., Steele, J.D., Lawrie, S.M., Frith, C.D., et al., 2006. Functional specialization within rostral prefrontal cortex (Area 10): a meta-analysis. J. Cogn. Neurosci. 18 (6), 932–948.

Hooker, C., Verosky, S.C., Germine, L.T., Knight, R.T., D'Esposito, M., 2008. Mentalizing about emotion and its relationship to empathy. Soc. Cogn. Affect. Neurosci. 3, 204–217.

Jenkins, C.A., Mitchell, J.P., 2010. Mentalizing under uncertainty: dissociated neural responses to ambiguous and unambiguous mental state inferences. Cereb. Cortex 20 (2), 404–410.

Leslie, A.M., Polizzi, P., 1998. Inhibitory processing in the false belief task: two conjectures. Dev. Sci. 1, 247–253.

Leslie, A.M., German, T.P., Polizzi, P., 2005. Belief desire reasoning as a process of selection. Cogn. Psychol. 50, 45–85.

Lieberman, M.D., 2007. Social cognitive neuroscience: a review of core processes. Annu. Rev. Psychol. 58, 259–289.

Mar, R.A., 2011. The neural bases of social cognition and story comprehension. Annu. Rev. Psychol. 62, 103–134.

Mars, R.B., Sallet, J., Schüffelgen, U., Jbabdi, S., Toni, I., Rushworth, M.F.S., 2011. Connectivity-based subdivisions of the human right "temporoparietal junction area":evidence for different areas participating in different cortical networks. Cereb. Cortex. http://dx.doi.org/10.1093/cercor/bhr26 (advanced online publication).

McCleery, J.P., Surtees, A., Graham, K.A., Apperly, I.A., 2011. Neural time-course of theory of mind. J. Neurosci. 31 (37), 12849–12854.

Mitchell, J., 2008. Activity in the right temporo-parietal junction is not selective for theory-of-mind. Cereb. Cortex 18, 262–271.

Mitchell, P., Lacohee, H., 1991. Children's early understanding of false belief. Cognition 39 (2), 107–127.

Perner, J., Lang, B., 1999. Development of theory of mind and executive control. Trends Cogn. Sci. 3, 337–344.

Rothmayr, C., Sodin, B., Hajak, G., Döhnel, K., Meinhardt, J., Sommer, M., 2011. Common and distinct neural networks for false belief reasoning and inhibitory control. Neuroimage 56, 1705–1713.

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

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

Saxe, R., Andrews-Hanna, J. (n.d.). Current Localiser stories. Retrieved July 19, 2009 from http://saxelab.mit.edu/stimuli.php.

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. Psychol. Sci. 17 (8), 692–699.

Saxe, R., Wexler, A., 2005. Making sense of another mind: the role of the right temporo-parietal junction. Neuropsychologia 43 (10), 1391–1399.

Saxe, R., Brett, M., Kanwisher, N., 2006a. Divide and conquer: a defense of functional localizers. Neuroimage 30, 1088–1096.

Saxe, R., Schulz, L., Jiang, Y., 2006b. Reading minds versus following rules: dissociating theory of mind and executive control in the brain. Soc. Neurosci. 1 (3–4), 284–298.

Scholz, J., Triantafyllou, C., Whitfield-Gabrieli, S., Brown, E.N., Saxe, R., 2009. Distinct regions of right temporo-parietal junction are selective for theory of mind and exogenous attention. PLoS One 4 (3), 1–7.

Sommer, M., Döhnel, K., Sodian, B., Meinhardt, J., Thoermer, C., Hajaka, G., 2007. Neural correlates of true and false belief reasoning. Neuroimage 35, 1378–1384.

van der Meer, L., Groenewold, N.A., Nolen, W.A., Pijnenborg, M., Aleman, A., 2011. Inhibit yourself and understand the other: neural basis of distinct processes underlying Theory of Mind. Neuroimage 56 (4), 2364–2374.

van Overwalle, F., 2009. Social cognition and the brain: a meta-analysis. Hum. Brain Mapp. 30, 829–858.

Vogeley, K., Bussfeld, P., Newen, A., Herrmann, S., Happe, F., Falkai, et al., 2001. Mindreading: neural mechanisms of theory of mind and self-perspective. Neuroimage 14, 170–181.

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

Wimmer, J., Perner, J., 1983. Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. Cognition 13, 103–128.

Young, L., Dodell-Feder, D., Saxe, R., 2010. What gets the attention of the temporo-parietal junction? An fMRI investigation of attention and Theory of mind. Neuropsychologia 48, 2658–2664.